THE DIGITAL MIND

How Science Is Redefining Humanity

数字思维

科学将如何重新定义人类、思维和智能？

[葡] 阿林多·奥利维拉（Arlindo Oliveira）_著　胡小锐 _译

中信出版集团

VISIT…

LANZAROTE Caliente.com

版权信息

书名:数字思维

作者:[葡]阿林多·奥利维拉

译者:胡小锐

出版时间：2020-01-01

ISBN:9787521711684

中信出版集团制作发行

[最新电子书免费分享社群，群主V信 1107308023 添加备注电子书]

献给

安娜·特蕾莎

推荐序

从物理形态上看，计算机毫无生气；而要把它作为一种能思考的机器来看的话，计算机又不堪重任。我们很久以前或许有过这样的念头，但这本书的作者从未这样想过。在几年以前，甚至是几十年前，我们就听说他经常会打破和谐的对话气氛，独树一帜地提出他自己的观点。他认为，有迹象表明推动生命系统进化的自然过程正在进驻新的数字领域；随着时间的推移，这些迹象变得越发明显，结果也变得越发奇妙。那些（像我一样）被他的独树一帜的想法吓了一跳的人，一直希望听到他说出更多的惊人言论。毕竟，阿林多是这方面的权威。他杰出的学术生涯跨越了从电路设计到机器学习的多个领域，而且越来越多地涉及生物学领域。因此，对于能思考的机器的发展前景，他有很多话要说，就不足为奇了。但是，正如这本书的读者将会发现的那样，它不仅呈现了惊人的言论，而且数量之多远超我们的想象！

作为加州大学伯克利分校计算机科学专业的一名学生，阿林多还辅修了神经科学。在某种程度上，我为他的远见卓识感到庆幸。在我看来，这为他后来的成就奠定了坚实的跨学科基础，但他却一反常态地将他辅修神经科学的举动自谦为年轻时“收效不大”的冲动。事实上，早在那时他就已经对数字思维产生了一些疑问，而且这些疑问永远不会消失。当时，他得到的关于认知的答案根本讲不通。正如你将在这本书中发现的那样，阿林多乐于推测，同时能给出合理的解释，并交代清楚历史背景。我指的不是在理论上可解释，而是指有线索可循，最终的原因都能归结到包含其中的电路上。作为一名生物学家，这种无法抗拒的冲动是我们熟悉的。对复杂过程的理解总是包含多个

层级的成分，从基因到细胞，最后一直到生态系统。同样的系统性原则引导着这次通往数字思维的旅程。光是巧妙地讲故事，肯定“收效不大”。

现在，我们发现一场数字风暴席卷了我们，我们的所有感知和行为都处于这场风暴的笼罩之下。新型设备以及为之提供支持的云服务平台限定了我们的交流方式，也限定了我们的活动范围。更重要的是，在我们更加专注地思考时，这种数字机器会大大拓展我们的视野。现在，神经科学通过先天和经验学习建立的模型来预测感觉输入，对大脑如何计算这个问题有了更深入的了解。正如人们预期的那样，递归是对数字机器的一个考验，而思考如何感知传感器同样会充分调动我们的思维。在涉及有意识的计算时，有机和无机之间的界限显然不再会引发分歧。说到有意识的计算，你可能会又一次以国际象棋或围棋为借口，寻找与阿林多进行这些对话的机会（不要抱太大希望！）。关于这些问题，他已经在计算领域探索和实践了几十年，他应该就此写一本书……

我本打算在最后一段为我的这篇推荐序画上一个完美的句号。但令人意想不到的是，阿林多在这本书的最后两章讨论了我们该何去何从的问题。在探讨了数字意识的可能含义和这个进程必然展现的前景这两个问题之后，这两章并未就此打住，而是继续深入下去。这本书是按时间顺序写作的，就像小说一样，所以在这里我不会剧透。犹如阅读任何一本好书，我们不仅要回答作者提出的问题，还要回答我们自己的问题。就我个人而言，我不禁认为数字生物和数字思维的出现，也是推动生物进化迈向更复杂结构的内衡驱力（homeostatic drive）的一部分。要构建数字机器并让它处理复杂世界中难以预测的变化，除了让它学会思考之外，还有什么其他更好的方法吗？正如这本书中解释的那样，思维可能是必然的结果，包括数字思维和其他思维。相反，物理学家可能更倾向于看到更多原子载体的信息属性的表征。无论如何，我们当前的数字化更像一颗种子，必将孕育出一个更加智慧的数字未来。

最后，我要提醒大家，如果你家里有十几岁的孩子，在把这本书推荐给家人之前，你自己一定要先读完。这本书可能会让我们感到不舒服，但它其实是为新生代写的。根据我的孩子的表现，他们可能会在这本书中找到一度困扰他们的问题的答案，即我们是如何让他们陷入数字化困境的。在这种情况下，他们会像我们一样困惑不解，不知道人类为什么对这个奇妙而不可避免的结果毫无察觉。换句话说，如果这本书到了孩子的手里，你可能就很难拿回来了。

乔纳斯·阿尔梅达

美国石溪大学生物信息学家

序言

看似毫不相干的计算机科学、生物学和神经科学一直吸引着我，并以这样或那样的方式影响着我的生活和职业选择。细心的读者会注意到，本书简直就是一个大拼盘，里面有关于这些学科的很多想法——我在过去30年中遇到的各种各样的想法。这些想法大多来自我结识的多位才华横溢的作家、科学家和哲学家，有的想法是他们与我面对面时提出的，但大多是通过他们的作品表达出来的。

本书在某种程度上似乎有些多余，因为要理解书中讨论的主题，我们可以直接搜索原始资料。尽管如此，我相信没有多少人拥有像我一样的人生经历，也没有多少人像我一样有幸去过那些迷人的地方、接触过那些有趣的思想家和概念。

我很早就迷恋上了数学和科学。我总想弄清楚各种事物的工作原理，而学习科学似乎能让我在养家糊口的同时，还可以做我喜欢的事。小时候的我对科学很痴迷，有好几年我都在折腾我的那些化学实验装置。当时“安全化学装置”的概念尚未提出，我每天都会想些新方法来合成化学物质，希望可以制造出炸药或其他危险的化合物。由于没有人指导我们的化学探索活动，所以烧瓶、试管、烧杯和其他化学仪器经常会出于破裂、崩解、爆炸或其他原因而变成一堆垃圾，但这个过程为我和我的朋友提供了无尽的乐趣。此外，对我们来说，这似乎就是科学。

之后几年里，我买了一些图书、杂志，尝试自己制作一些电子产品。第一次组装电子电路时，我因为一无所知而吃尽苦头。我买回来

一个个电子元件，然后照着示意图一步步组装。画这些示意图的人都是行家，而一窍不通的我根本看不懂。

20世纪70年代末，市面上出现了非常原始版本的个人电脑。在买了几台比较原始的机器之后，我自豪地拥有了一台辛克莱ZX Spectrum计算机。这种小型个人电脑拥有48千字节的存储器和集成键盘，堪称当时的一个技术奇迹。它用一台电视机作为显示器，可以用Basic（培基，一种直译式程序设计语言）编程。它没有永久性存储器，新程序都必须保存在标准磁带录音机使用的磁带上。通常情况下，由于磁带和录音机的质量问题，我保存的程序都无法再次调用，但我仍然会投入大量时间编写程序。ZX Spectrum帮助我迈出了编程的第一步，我也自此与计算机结下了不解之缘。

在葡萄牙里斯本高等理工学院学完电气工程后，我专攻数字电路设计，这个领域当时还处于起步阶段，与许多其他通信和信息技术并行发展。具体来说，我从事的是超大规模集成电路（VLSI）的相关研究。超大规模集成电路是一个通用术语，指包含很多设备的集成电路。在极具影响力的科学家、政治家何塞·马里亚诺·加戈的主持下，葡萄牙与欧洲核子研究组织（CERN）签署了一项协议。因为这项协议，我得到了一个在日内瓦的暑期实习机会。在欧洲核子研究组织实习期间，我借助控制粒子物理实验中的高能粒子束的设备，测试了我的编程技能，并感受到物理学界对寻找支配宇宙的法则充满了热情。在日内瓦的那个夏天，我对粒子物理学的兴趣更加浓厚了。

在欧洲核子研究组织的实习结束之后，我觉得有必要进一步了解正在快速发展的数字电路领域的动向。于是，我向加州大学伯克利分校提出入学申请并被录取。这所大学擅长开发设计集成电路所需的新工具、电路和技术。在那里，我进入了集成电路设计的技术新天地，其中有很多新技术都是CAD（计算机辅助设计）小组开发的，他们当时的负责人是阿尔贝托·圣乔瓦尼－文森泰利（我的导师）、罗伯特·布雷顿和理查德·纽顿。加州大学伯克利分校在很多领域都不乏顶尖

人才，这让我有机会去了解我不熟悉的领域。其中有两个领域吸引了我，并明确了我未来的学习方向——算法和神经科学。

我在加州大学伯克利分校跟着理查德·卡普学习算法，他是算法和复杂性理论的创建者之一。建立可用于设计集成电路的算法是加州大学伯克利分校CAD小组的任务，集成电路的设计人员使用算法去创建、验证、模拟、放置和连接晶体管，晶体管是所有电子器件的工作单元。如果没有这些复杂的算法，今天的集成电路就无法设计和制造出来。

在加州大学伯克利分校，我也有机会了解到另一个即将与算法和计算机产生密切联系的领域——神经科学。我辅修了神经科学，努力掌握了一些进化、神经元和大脑方面的知识，但这足以让神经科学研究成为我的另一大爱好。当时，一些计算机科学家已经着手开发处理生物数据的算法，这个领域后来被称为生物信息学。基于我的兴趣，开发解决生物学问题的算法对我而言是一个显而易见的选择。多年来，我在这个领域做出了一些贡献。

最近，我有幸评审了人类脑计划（Human Brain Project）的一些提案。这些提案涵盖脑模拟、生物信息学和脑治疗等多个领域，让我越发意识到这些领域之间的联系的重要性。人类脑计划的目标是，使用计算机来支持脑研究。从某种程度上说，随着计算机在理解人脑这一核心问题上的应用，我的科学之旅绕了一圈后又回到了原点。

这趟旅程受到了很多人的影响。我的导师、同事和学生陪伴着我，与我一起工作，并在我取得的几个相关科学成果中发挥了重要作用。曾经和我一起工作的同事现在遍布世界各地，与他们的交流对我产生了无法用语言描述的影响。

不过，我认为在这几十年里对我影响最大的可能还是我读过的那些书，其中有几本深刻地改变了我对世界的认知，如果没有它们，我的生活肯定会与现在大不相同。当然，如果没有在我之前的许多作者

对计算、进化、智能和意识之间关系的探索，就不可能有你们现在读的这本书。

对我产生过影响的人有很多，其中有一些作者值得我在这里列出他们的名字。理查德·道金斯和史蒂芬·杰伊·古尔德让我见识到进化的奇迹。丹尼尔·丹尼特和侯世达是我从事人工智能研究的领路人，激发了我对大脑、思维和意识问题的兴趣。戴维·休伯尔清晰地描述了大脑的某些区域的组织方式，让人觉得只要多付出一点儿努力，就可以解释整个大脑的行为。史蒂芬·平克、马文·明斯基和罗杰·彭罗斯深刻地影响了我对人类思维的运转方式和智能的含义的看法，尽管在某些方面我不同意他们的观点。凯文·凯利对技术前景的想法，以及雷·库兹韦尔对技术奇点的坚定信念，也对我产生了强烈的影响。詹姆斯·沃森和约翰·克雷格·温特为我提供了关于生物科学过去50年的发展历程的独到见解。埃里克·德雷克斯勒对纳米技术前景的展望，改变了我对“非常小”的概念和对世界的看法。承现峻、奥拉夫·斯波恩斯和史蒂文·罗斯对他们参与的旨在理解大脑的项目描述，使我受到极大的启发和鼓舞。尼克·博斯特罗姆解决了本书涉及的许多问题，并且以我做梦也想不到的方式做了进一步拓展。贾雷德·戴蒙德和尤瓦尔·赫拉利深刻改变了我看待人类历史和人类进化的方式。

科幻小说也对我产生了强烈影响，我是一个狂热的科幻小说读者。艾萨克·阿西莫夫、亚瑟·克拉克和罗伯特·海因莱因，让我对科幻小说产生了持久的兴趣。格雷格·伊根、弗诺·文奇、尼尔·斯蒂芬森、拉里·尼文和查尔斯·斯特罗斯对未来的展望如此清晰且富有挑战性，以至于人们很容易相信它们总有一天会发生，尽管它们奇怪又令人不安。

我竭力让所有对本书谈论的话题感兴趣的人，都能轻松地读懂它。本书的目标读者是对科学和技术感兴趣的普通人，不需要事先对

本书涵盖的诸多领域的知识有所了解。最后三章不包含任何技术相关内容，它们呈现了本书的中心论点，可以独立阅读。

缺乏技术背景的读者不必因为偶尔出现的方程式、图表和数学论证而沮丧。我会尽可能地提供充足的信息来帮助你们掌握一些技术细节，但在大多数情况下，你们可以通过这些附带的解释来掌握本书的核心思想，就算略过细节也不会遗漏任何重要信息。

第1章

快速奔跑的红皇后

对一个漫不经心的观察者来说，计算机、细胞和大脑似乎没有什么共同之处。计算机是人类为简化和改善他们的生活而设计的电子设备；细胞是所有生物的基本要素，是通过进化产生的生物实体；大脑是我们思想的容器和创造者，里面装着希望、恐惧和欲望。

无论如何，计算机、细胞和大脑都是信息处理设备。计算机代表最新的数字形式的信息处理方式，而在此之前，信息处理是由生物完成的，实际上就是在混沌中创建出秩序。计算机、细胞和大脑都是同宇宙一样古老的复杂物理和化学过程的结果，是进化的最新产物，是始于亿万年前的一场竞赛的胜利者。

万事万物皆相互关联

我们都是由原子构成的，原子或者产生于140亿年前标志着时间开始的大爆炸，或者产生于大爆炸后几十亿年内的遥远恒星爆发。我们的每一次呼吸都与恺撒大帝的最后一次呼吸一样，吸入的是相同的氧分子，它们被反复使用了成百上千年，维系着地球上的所有生命。我们现在知道，蝴蝶扇动翅膀确实可能会影响好几个月后的一场飓风，而一个物种能否存续，可能在很大程度上取决于地球上所有其他物种的存续。始于大约40亿年前的进化过程造就了我们，也造就了我们周围的一切，包括我们拥有和使用的许多设备及工具。

本书涉及许多个在我看来相互之间存在强联系的不同领域，包括计算机、进化、生命、大脑、思维，甚至还有一点儿物理学。你可能认为这些领域既无交集，也不相关。我将努力向你们证明，它们之间是紧密相连的，有一条共同的线把物理学、计算和生命联系在一起。我知道，本书每一章的主题都可以写成一本专著（有的主题甚至可以写成好几本专著）。对于所有这些领域，我只能点到即止，帮助读者理解其基本原理，但我希望本书的连贯性不会因此受到影响。

从某种程度上说，世间万物皆可归结为物理学。欧内斯特·卢瑟福说过，“除了物理学，其他所有科学都是集邮”，意思是物理学定律应该足以解释所有现象，而所有其他科学不过是物理学定律的不同抽象形式。但是，物理学不能用于直接解释或研究世间万物。要理解宇宙是如何运行的，就离不开计算、生物学、化学和其他学科。不过，这些学科并非毫无关联，适用于其中一个学科的原理也适用于其他学科。[最新电子书免费分享社群，群主V信 1107308023 添加备注电子书]

我们之所以意识到一切事物都是相互关联的，在很大程度上是因为我们对世界的理解和我们掌握技术的能力得到了提升。直到最近，科学的进步才让我们理解了这样一个事实：我们接收到的日光、磁体的行为和大脑的工作原理，都可以用相同的方程来描述。仅仅几百年之前，这些还都是互不相干的现实——对被无知蒙蔽了双眼的人类来说，这些神秘事物之间毫无关联。

科学不断提升我们对世界的理解，借着这个东风，几千年来技术以越来越快的速度不断改变着我们的生活，但“技术”（technology）一词的出现时间并不长。1772年，德国学者约翰·贝克曼创造了这个词（其本意是“工艺科学”），并在某种程度上创造了这个概念。在贝克曼创造这个词之前，技术的不同方面分别被称为工具、艺术和工艺。贝克曼在若干著作中都使用了“技术”一词，包括后来被翻译成英文的《技术或工艺、工厂和制造商知识指南》（*Guide to Technology, or to the Knowledge of Crafts,Factories and Manufactories*）。但是，在20世纪以前，“技术”一词很少在通用语中出现。20世纪下半叶，这个词在通用语中的使用频率稳步增加。今天，它是政治、社会和经济文本中的一个十分常见的词语。

随着技术的发展，人们认为创新和新颖的事物是文明的内在组分。技术、社会和经济的不断变化在我们的日常生活中根深蒂固，以至于我们很难理解并非如此的古代。几百年前，变化非常缓慢，以致大多数人都预期未来与过去区别不大。所以，少有人甚至没有人持有未来将改善人们生活的观念。当变化频繁地发生，人们不仅可以察觉到，而且可以预期变化时，一切都不一样了。随着技术的出现，人们开始期望未来会带来可以改善他们日常生活的新事物。然而，我们中的许多人现在却担心这些变化来得太快，而且影响过于深远，以至于普通人可能无法理解它们。

老年人很难适应新事物

如果你已经几十岁了，那么你可能会觉得日新月异的技术让你目不暇接，层出不穷的新设备、新时尚让你眼花缭乱，你几乎跟不上变化的步伐。现在年轻人做的事情和用的东西对老年人来说越来越陌生，新的潮流、工具和玩具让老年人力不从心。小孩子比你更会用智能手机，打起电脑来更得心应手，会玩你不懂的游戏，还会一窝蜂地访问你不感兴趣的新网站。他们甚至不知道VCR（盒式录像机）是什么，在20世纪80年代，VCR可是一项堪称奇迹的技术，不过也是出了名的难操作。即使是CD（小型光盘）和DVD（数字光盘）这两项20世纪最后几十年才出现的数字技术，今天也快要退出历史舞台了（它们存在的时间比被它们取代的黑胶唱片和胶卷明显短得多，后两者存在了近100年）。

但是，如果你出生于21世纪，那么你会理解不了你的父辈和祖辈为什么难以适应技术创新。遇到新想法和新奇的小玩意儿时，你会轻而易举地接受它们；面对最新的应用程序、网站或社交网络时，你也会觉得驾轻就熟。

不过，我们这代人中那些受过良好教育，甚至技术纯熟的人，根本不会拒绝新技术。我们这代人发明了计算机、手机、互联网和DNA（脱氧核糖核酸）测序等许多东西。我们应该能够理解和使用技术在未来几十年里可能抛给我们的任何新事物。

我相信，技术将以日益增长的速度不断变化。就像我们的父辈和祖辈在我们之前经历的那样，我们掌握的技术知识很可能会迅速过时，而且我们难以理解、使用和跟上未来几十年的技术发展。俗话说得好：“老年人很难适应新事物。”

我们的子孙后代是否也会不可避免地落伍于时代，这将取决于技术持续改变的态势。新技术的发展会越来越快吗？抑或是我们身处技术发展的黄金时代，事物正在以前所未有的速度发生变化？

8岁的时候，我喜欢去看望我的祖父。他住在一个小村庄里，在当时的我看来，那是一个相当偏远的地方，距离葡萄牙的首都里斯本大约60英里[注]。因为道路蜿蜒崎岖，从我家开车到那里要花上半天的时间。祖父经常驾着一辆小马车，载着我一起去田里干活。对一个小男孩来说，坐马车是一件非常兴奋的事，因为它给人一种回到过去的感觉。另外，祖父家没有电灯、冰箱、电视，也没有书，这些强化了我回到过去的感觉。因为我的祖父母都过着旧式的生活，他们认为这些“现代”技术都不是必需品。

相比之下，我的父母结婚后就离开了这个小村庄，他们更加熟悉先进的技术。他们受过教育，拥有一台电视机，甚至还买了一辆汽车。而对我的祖父母来说，这些技术毫无意义。他们对电视、报纸或书籍不感兴趣，因为里面报道或提及的大多数现实都与他们的日常生活没有任何关系。汽车、火车和飞机对那些很少走出他们的小村庄，也从未想过要走出去的人来说，同样毫无意义。

40年后，我的父母仍然拥有一台电视机，仍然读书，仍然开着车四处兜风。我本来以为，他们那代人和我们这代人之间的技术差距会大大缩小。我本来想象，40年后，在理解技术的能力方面，他们那代人与我们这代人之间的差距将会远小于他们与他们的上一代人之间的差距。然而，事实并非如此，技术差距似乎在逐代增加。我的父母现在已经80多岁了，他们从未意识到计算机是一种通用机器，可以用来玩游戏、获取信息或进行交流；他们也从不理解计算机只是复杂的信息设备网络的一个终端，可以随时随地根据人们的需要有针对性地传输信息。他们中的许多人都不知道这样一个事实：除了技术限制造成的小小不便（很快就会消失）以外，计算机完全可以取代书籍、报纸、广播、电视等几乎所有用于传递信息的设备。

你可能认为我们这代人了解计算机的功能，不会轻易地落后于技术发展。毕竟，我们中的几乎所有人都明确地知道计算机是什么，很多人甚至知道计算机的工作原理。这些知识应该能给我们一些信心，让我们认为自己不会像我们的父辈和祖辈那样，被技术的新发展超越。不过，这种自信可能是错误的，这主要是因为红皇后效应（Red Queen effect）。这个概念之所以得名，是受刘易斯·卡罗尔的名著《爱丽丝镜中奇遇记》（*Through the Looking Glass*）中的一句话的启发："现在，你明白了吧，想要待在原地，你就必须竭尽全力地奔跑。"

红皇后效应源于这样一个事实：随着进化的持续进行，生物必须变得越来越复杂，其目的不是为了获得竞争优势，而只是为了能像系统中的其他生物一样存活下来，并变得更具竞争力。

尽管红皇后效应与进化和物种间的竞争有关，但它同样适用于竞争导致快速变化的其他所有环境，例如，商业或者技术。

我预计未来的每一代人落后于时代的程度都将远远超过他们的上一代人。30年后，我们对那时科技的了解程度甚至比不上我们的父母对今日科技的了解程度。我相信，世代淘汰的进程将不可避免地逐代继续加速，100年后，即使最基本的日常生活概念都会让我们这代人感到新奇。

克拉克第三定律指出，任何非常先进的技术都与魔法无异。在现代人眼中，100年后的科技可能非常新奇，看起来就像魔法。

从计算机和算法到细胞和神经元

新技术的出现并不是新鲜事。然而，在如此短的时间内出现了如此多的技术创新，这是前所未有的。在未来几十年里，我们将继续看到若干技术的迅速发展和融合，而在不久之前，它们还被视为互不相干。

其中一项是计算机技术，它是随着电子学和计算机科学的发展而出现的，尽管出现时间不长，但已经极大地改变了世界。计算机现在几乎随处可见，以至于对我们许多人来说，很难想象没有计算机的世界会是什么样子。然而，计算机之所以有用，只是因为它们可以执行算法。算法无处不在，它们是计算机存在的最终原因。如果没有算法，计算机将毫无用处。

算法开发人员的工作就是寻找最有效的途径，告诉计算机如何高效、正确地执行特定计算。算法是非常详细的“食谱”——计算机执行算法给出的一系列具体步骤，就可以获得某个特定的结果。我们在学校都学过的加法运算，就是一个广为人知的算法案例。无论两个数有多大，只要我们按照这种算法给出的一个个具体步骤，就可以求出这两个数的和。加法运算是现代计算机的核心，计算机的所有应用都离不开它。

算法是通过某种特定的程序设计语言传递给计算机的。算法本身不会随着程序设计语言的变化而改变，它们只是描述如何达到某个结果的抽象指令序列。在我看来，算法设计是数学和计算机科学中最优雅、最迷人的领域之一。

算法往往是为了特定的目的而开发出来的。算法有很多应用领域，其中有两个将在未来的技术发展中发挥基础性作用。

第一个应用领域是机器学习。机器学习算法可以赋予计算机从经验中学习的能力。你可能认为计算机不会学习，而只会按照明确的指令运行，但事实并非如此。计算机的学习方式有很多种，当我们看电视、上网、使用信用卡或打电话时，我们利用的就是计算机的学习能力。在许多情况下，尽管我们无法看到实际的学习机制，但学习仍然发生了。

第二个应用领域是生物信息学，即应用算法来理解生物系统。生物信息学（亦称计算生物学）利用算法处理通过现代技术获得的生物学和医学数据。我们可以完成基因组测序，可以收集并利用有关生物机制的数据去理解生物系统的运行方式，这些在很大程度上是因为我们使用了特定的算法。生物信息学是一种能够建模和理解细胞及生物行为的技术，生物学领域取得的最新进展与生物信息学的发展有着密不可分的联系。

创造了所有生物的进化过程，在某种程度上也是一个算法。进化是在一个完全相同的平台上进行的，并且已经运行了大约40亿年，但它本质上仍是一个以生物繁殖能力最优化为目标的算法。40亿年的进化不仅创造了细胞和生物，还创造了大脑和智能生物。

尽管进化创造了各种各样的生物，各种各样的新技术也被发明出来，但人类一直是地球上高级智能的独一无二的拥有者。正是这种特性使得人类能够统治地球，并根据他们的需要和欲望来改造地球，有时还以牺牲其他重要因素为代价。

然而，技术的进化如此迅速，以至于我们现在第一次面临其他实体——我们创造的实体——真的有可能拥有智能的情况。这种可能性产生于过去50年里发生的计算机技术（包括人工智能和机器学习）革命，也产生于我们在理解生物（尤其是人体和人脑）方面取得的重大进步。计算机技术的历史只有几十年，但它已经改变了我们日常生活

中的许多事物，如果没有计算机，那么我们现在看到的文明根本不可能出现。物理学和生物学也取得了巨大进步，我们第一次拥有了可用于详细了解人体和人脑工作原理的工具和知识。

今天，医疗技术的进步已经大大增加了大多数人的预期寿命，但在未来几十年里，我们控制或治愈致命疾病的能力有望实现前所未有的提升。原因在于，生物学、物理学和计算领域的新技术，有助于我们不断增进对生物过程的理解。最终，我们将实现对生物过程的充分理解，从而在计算机上复制和模拟这些过程，为医学和工程学开创新的可能性。

随着医学和计算技术的进步，我们终有一天会充分了解大脑的工作原理，并尝试用数字支持的形式复制智能，也就是说，我们可以编写一个由数字计算机来执行的程序，它会展现出智能。在本书中，我们将探索这种可能性。有若干种方法可以实现这个目标，但它们都会造成非生物思维产生并成为我们社会中一员的局面。我称它们为“数字思维”（digital mind），原因在于，几乎可以肯定的是，数字计算机技术的存在将使它们成为可能。非生物思维可能很快就会出现在地球上，这一事实将引发一场迄今为止我们从未见过的社会革命。然而，大多数人不仅对这场革命的可能性视而不见，也看不到它将给我们的社会和政治体系带来的深刻变化。本书的另一个目的是，增加公众对这场革命的影响力的认知。

我在书中做出的大多数预测可能都是错误的。到目前为止，未来总能创造出超越人类想象力的东西，更奇怪也更具创新性和挑战性。接下来自然也不例外，如果我有幸能活到那一天，我肯定会因为那些出乎我意料的新技术、新发现而惊讶不已。不过，这就是技术的本质，做预测本就不容易，预测未来更是难上加难。

是时候开启我们的探索之旅了，它的起点是技术的起源，终点是思维的未来。我要告诉大家的第一件事是，指数增长是生命宏图中的一个内置模式，也是许多技术发展的特征之一。

1. 1英里≈ 1.6千米。——编者注

第2章

技术的指数性质



自人类诞生以来，文化和文明带来的多种多样的技术，已经深刻地改变了人们的生活方式。每种技术甫一问世，就会改变个体、部落和整个群体的生活。然而，变化的速度并不是恒定的。自从几十万年前第一次创新活动发生以来，新技术的开发和应用速度一直在增长。[最新电子书免费分享社群，群主V信 1107308023 添加备注电子书]

技术创新几乎和人类的历史一样古老，甚至早在300多万年前，一些非人类祖先可能已经使用过基础技术。不过，新技术的引入速度并非一成不变。在史前时代，一项新技术的引入往往要花几万年甚至是

几十万年。但近1 000年来，新技术的出现速度要快得多。在20世纪，新技术每隔几年就会出现，此后速度便不断加快。

新技术总是在已有技术的基础之上发展产生，这是技术的本质，因此，每一个重要的新发展都比之前的那些花费的时间和精力更少。这也使得技术变革的速度越来越快，因此有许多人认为，随着重要的新技术被开发出来并取代旧技术成为变革的引擎，技术变革的速度将在很长一段时间内呈指数增长。

史前技术

虽然若干物种曾多次开发并使用石制和木制工具，但我们可以认为人类完成的第一个重大技术创新是火的发现（更准确地说，是火的控制使用）。尽管这一重大事件的确切发生日期尚存争议，但它可能发生在旧石器时代的中后期，也就是100多万年前到40万年前，早于智人的出现时间。

火的控制使用显著改变了史前人类的生活习惯。之后的几代人逐渐习惯了这样一种观念：烤熟的肉吃起来更容易，煮熟的蔬菜味道更佳。蒸煮使食物更易于消化，人们也能从食物供应中获取更多的卡路里。火的控制使用可能是人类大脑尺寸增加的主要原因。

从我们的角度看，在人类驯化动植物之前的几千年里，似乎没有发生什么大事。当然，事实并非如此。其间发生了一些重大的文化变革，包括尤瓦尔·赫拉利所说的认知革命，它使人类学会用语言来交流抽象的思想和观点。在农业革命前的几千年里，发生了许多重大的技术突破。弓箭、针线、斧头和矛都不止一次地被发明出来，并改变了当时的社会。这些发明或发现的确切日期大多无法确定，但它们可能都发生在5万年前至1万年前。每1万年（大约400代）出现一项重大发明，这样的变化速度似乎不太激动人心，但我们必须记住，自我们的祖先从树上下来到他们的生活习惯发生重大变化，这个过程长达几十万年。

约1.1万年前到1万年前，人类已经散布到大多数大陆上，并开始了乡村生活。几千年后，许多文明开始驯化动植物。这被视为继火的控制使用之后的第二次重大技术革命，通常被称为农业革命，它深刻地改变了人们的生活和互动方式。从狩猎采集的游牧生活方式向以农

业为基础的定居生活方式的转变，不一定意味着人们的生活过得更好了。随着人口的增长，人类越来越依赖于固定的食物供应，更大的群落导致传染病发病率不断上升，也带来了争夺霸权的斗争。尽管如此，在大多数地方，这些变化的发生肯定历经了许多代人，而且，它们很可能并未引起个体生活方式的深刻变化。

从那时起，技术进化的增长速度就从未停止。尽管确切的日期尚存争议，但最早的轮子可能出现在公元前3 500年的美索不达米亚。让人惊讶不已的是，美洲三大文明——阿兹特克、玛雅和印加文明——都没有用过轮子，这可能是因为他们缺乏便利的耕畜。不过，绝大多数的东西方文明都广泛地使用了这种精妙的工具。

在过去的1万年里，其他许多旨在节省体力的发明都出现了。在技术进步方面，所有文明或世界各地并非步调一致。尽管确切的原因尚存争议，但有一个理由毋庸置疑，即技术的不断进步加剧了技术发展之初的微小时间差，并导致技术发展水平迥然不同。相对小的环境差异，比如，耕畜的可获得性，来往于不同地方之间的难易程度，以及可以栽种的作物类型，造成技术（比如农艺、运输工具和武器）的引入时间不同。这些反过来又导致更先进的技术在不同文明中的出现时间有先有后，而更先进的技术发展与更基础的技术有关，尽管有时这种相关性不太明显。与更基础的技术相关的技术发展，推动了以信息为基础的技术发展（比如计数和书写），以及其他社会进步。最终，那些领先一步的文明占据了优势，使西方社会和文明获得了支配地位，这在很大程度上得益于其出类拔萃的技术能力。不难看出，在技术竞赛中起步较晚的文明处于明显的劣势地位。技术发展的速度或时间上的微小差异后来会演变成巨大的差异。除极少数特例之外，西方文明凌驾于整个已知世界之上，而其他发达文明则被摧毁、吸收或改造。

第一次和第二次工业革命

18世纪晚期的工业革命是技术发展史上最广为人知的一个间断点。工业、农业、采矿业和交通运输业的技术革新，使英国的社会结构发生了重大变化，后来又在欧洲、北美洲和世界其他地区引发了同样的变化。这次革命是人类历史上的一个重要转折点，几乎改变了人们日常生活的方方面面，也大幅加快了经济发展。在过去的几千年里，世界经济的增速每过900年左右就会加倍。在这次技术革命之后，经济增速加快，现在经济产出每过15年就会加倍。

这次工业革命有时会被划分为两次工业革命，第一次革命（发生在18世纪晚期）以纺织业生产方式的变化为标志，第二次革命（开始于19世纪中叶）的标志则是蒸汽机以及其他交通运输和通信技术的发展，其中很多技术都以电力为基础。这两次革命深刻地改变了消费品的生产方式，并使电话、火车、汽车、飞机等通信和交通运输技术得到了广泛应用。

第一次工业革命始于英国，然后迅速蔓延到美国以及西欧和北美的其他国家。它给社会带来的深刻变化，与纺织、冶金、交通运输和能源等领域中发生的为数不多的技术创新密切相关。

在第一次工业革命引发纺织行业巨变之前，纺纱和织布大多是在织工的小作坊里完成的，主要满足国内的消费需求。在家工作的织工既纺纱又织布，他们与商人签订合同，原材料大多由商人提供。纺纱和织布的技术变革带来了生产率的大幅提高，在工业革命中起到了重要作用。

包括珍妮纺纱机和走锭纺纱机在内的一些机器取代了纺纱工人，使纱线的生产效率更高，成本也更低廉。在工业革命之前织布一直是一种手工艺，但在机械织布机发明（其中包括飞梭等许多创新）后实现了自动化，这大大提高了纺织业的效率。随后，提花织机的发明又大大简化了复杂图案纺织品的生产过程。提花织机与我们当前讨论的内容尤其相关，因为它使用穿孔卡片来控制织物的颜色。19世纪后期，查尔斯·巴贝奇利用穿孔卡片，开发出第一台机械式通用计算机。

第一次工业革命以纺织业和冶金业为中心，而第二次工业革命的特点是：铁路的大量修建，机器在制造业的广泛使用，蒸汽动力的应用增加，石油开始投入使用，电的发现和利用，电信业的发展，等等。交通运输技术的变革以及用于制造、分销产品与服务的新技术，掀起了第一次全球化浪潮，这一现象为不同经济、文化和生态的发展提供了源源不断的动力。

不过，在电及其大量应用带来的变化面前，即使是改良的交通运输技术也会相形见绌。基于电的技术发展是今天的互联世界的源头，新闻和信息以光速传播，创建出第一个真正意义上的全球社区。

在前两次工业革命期间，人们发起了多次反对引入新技术的运动，理由是新技术会摧毁就业机会并取代工人。其中最著名的是勒德派运动，勒德派是19世纪初成立的英国工人组织，他们袭击工厂，毁坏设备，以抗议引进机械针织机、细纱机、动力织布机等新技术。他们所谓的领导人“内德·勒德”可能是虚构出来的人物，但“勒德派”一词却沿用至今，指代反对技术变革的人。

第三次工业革命

始于1970年前后的信息化时代通常被视为第三次工业革命。由于这场革命方兴未艾，因此我们还无法感受到它的全面影响。继第三次工业革命之后是否会有一个泾渭分明的第四次工业革命，仍是一个悬而未决的问题。事实上，为这个问题提供答案是本书的目标之一。

本书特别感兴趣的是与信息处理有关的技术发展。我认为，信息处理技术最终将超过几乎所有其他的现存技术，并在不远的将来取代那些技术。原因在于，信息处理以及记录、存储和传播知识的能力，是大多数人类活动的源头，它将变得越来越重要。

人类对处理信息的首要需求，可能源于他们准确保存储备物资和农业产出数据的需要。对狩猎采集者来说，几乎没有必要把信息记录下来并传递给其他人或后代。随着复杂社会的演变，书写成为一种必需品，人们需要记录缴税时应该上缴多少只羊，或者记录某人拥有多少土地。此外，存储和传播维持复杂社会运转所需的详尽社会准则，也使书写变得越来越有必要。

与许多其他技术不同，创造书面语言是一项非常复杂的任务，其独立进化可能发生了若干次。想要发明一种书面语言，人们必须弄清楚如何将句子或想法分解成小的单元，商定一套编写这些单元的统一系统，并确保第三方能理解全部内容，从而实现这种语言的全部价值。虽然其他独立发展的文字系统（包括中国和埃及的文字系统）可能早已出现，但公认最早的是公元前3500年至公元前3000年美索不达米亚平原的苏美尔人发明的楔形文字。许多其他的书写系统都是借鉴这些原始文字发明者的想法发展而来的。

现在，人们认为文字系统的最早用途是，记录应交付的或已生产的农产品数量。后来，书写进化为可以表示语言中的一般词汇，于是它被用来记录故事、历史，甚至是法律。《汉谟拉比法典》是古代美索不达米亚的一套著名的早期法律，由古巴比伦国王汉谟拉比颁布，可以追溯到公元前18世纪。它是迄今为止破译的最古老、最复杂的著作之一，其部分副本存在于石碑和各种泥板上。法典本身由数百条法令组成，描述了不同社会地位的人违反特定规则时应采取的恰当行动。

书写是信息处理技术上的一个重要发展，因为它第一次使远距离传播知识成为一种可能。更重要的是，书面记录的保存时间长于人类的记忆，甚至长于书写者的寿命。一条信息可以被保存下来，被其他人完善，并且易于复制、分发，这是前所未有的创新。

统计、记录和处理数量的相关能力是书写的源头，并与书面语言并行发展，从而形成了数学这个极其重要的产物。已知最早的数学应用是为了满足农业、商业和工业的实际需要而产生的。在公元前两三千年的埃及和美索不达米亚，数学被用来测算数量。类似的发展也发生在印度、中国和其他地方。早期数学具有经验性，以简单的算术为基础。

希腊人在认识到数学公理的必要性之后，发明了几何学、三角学、演绎系统和算术，这是数学最早取得的有记载的重大发展。许多著名的希腊人都对数学的发展做出了贡献，其中，泰勒斯、毕达哥拉斯、柏拉图、亚里士多德、希波克拉底和欧几里得，对我们今天熟悉的许多概念的建立起到了关键性作用。中国人和阿拉伯人继承了希腊人留下的数学遗产，并实现了许多重要的发展。此外，阿拉伯人还保存、翻译和增补了希腊人的著作。在今天被称作巴格达的地方，当时的一位重要数学家花拉子密引入了印度－阿拉伯数字和代数的概念。“算法”的英文单词algorithm就是从花拉子密的名字衍生而来的。

之后，数学领域取得的发展数不胜数，而且非常复杂，我在这里就不一一列举了。数学从简单的算术概念发展到古希腊人提出的几何学、代数和三角学概念，这个过程花费了几千年的时间，但它建立起现代数学这个惊人的庞大体系却只用了不到500年的时间。

关于数学，特别值得一提的是计算理论。计算是按照由一系列运算描述的明确模型进行某种运算的过程。计算可以借助模拟或数字设备来进行。在模拟计算中，一些物理量（比如位移、重量或液体体积）被用于给研究对象建模。在数字计算中，数字或符号被用于研究对象的离散表示，以便得到期望的结果。

几千年来，人们已经开发了许多执行模拟计算的设备，其中最引人注目的安提凯希拉装置可以追溯到公元前150年到公元前100年之间。它是源于希腊的一种复杂的模拟计算机，使用一组复杂的联锁齿轮来计算太阳、月亮，也许还有其他天体的位置。模拟计算有着悠久的传统，喜帕恰斯（约公元前190—前120年）发明的星盘，以及17世纪威廉·奥特雷德基于约翰·纳皮尔的对数概念发明的计算尺，都属于这个范畴。计算尺是一种非常有用和有效的工具，至今仍被许多工程师使用，本书的读者中可能也有人在使用它。

20世纪，许多模拟计算机被设计出来并投入应用。其中的两个代表性案例是美国海军安装在很多舰艇上的Mark I火控计算机，以及威廉·菲利普斯在1949年建造并用于为英国经济建模的MONIAC液压计算机。但是，模拟计算机在灵活性方面受到许多限制，这主要是因为每台模拟计算机都是针对特定的应用构思和建造的。通用模拟计算机在概念上是可能的（计算尺就是一个很好的例子），但总的来说，模拟计算机是为执行特定任务设计的，不能用于其他任务。

相反，数字计算机是通用计算机。只要简单地修改一下数字计算机执行的程序，它就可以执行各种任务。虽然辅助算术运算的复杂工具（比如，公元前500年左右出现的算盘）已经存在了数千年之久，但全自动计算的概念直到最近才出现。托马斯·霍布斯可能是第一个明

确提出“人脑的运行是机械运动”观点的人，他认为大脑就是一台计算机：

> 一个人在思考时，他的大脑想的就是一堆数的总和，或者一个数减去另一个数的差……这些运算不仅与数字有关，而且与所有可以相加和相减的事物有关。

但是，直到19世纪，查尔斯·巴贝奇才设计出第一台通用数字计算机——分析机，与他同时代的阿达·洛芙莱斯在此基础上迈出了至关重要的一步，即认识到数字计算机的作用不只是“鼓捣数字”。

指数趋势的惊人性质

在几乎所有技术领域中普遍存在的一个现象是：新技术的推出速度不断加快。几千年来，技术进步一直很缓慢，许多代人都过着与他们的父辈、祖辈差不多的生活；然后，一些新技术出现了，它们与下一次技术发展之间的时间跨度也比以前短得多。达到一个新阈值所花的时间远远少于达到前一个阈值所花的时间，这是数学领域的指数函数的特征。[最新电子书免费分享社群，群主V信 1107308023 添加备注电子书]

在数学中，指数函数是a^n形式的函数。因为$a^{n+1}=a\times a^n$，所以函数在点n+1处的值等于函数在点n处的值与指数函数的底数a的乘积。当n=0时，函数值为1，因为任何数（0除外）的0次幂都等于1。

本书感兴趣的是底数a大于1的情况。a可以远大于1，也可以只比1大一点儿。在前一种情况下，指数将会增长得非常快；即使n很小，函数值也会迅速变得很大。例如，即使n很小，函数10^n也会增长得很快。这比较直观，我们不应该感到惊讶。不太直观的是，即使底数很小，比如a为2或者1.1，a^n形式的函数也会增长得非常快。

我将介绍两种类型的指数增长，它们都与上述讨论有关。第一种类型与函数随时间呈指数增长的演变相关，第二种类型与几个相互独立的简单可能性组合之后引发的组合爆炸相关。

我要举的第一个例子是，象棋发明者与中国皇帝的那个众所周知的历史故事，以此说明指数趋势的惊人性质。传说在象棋被发明之后不久，酷爱这个游戏的中国皇帝就把发明者传唤到朝廷上并告诉他，作为世界上最强大国家的皇帝，可以答应他的任何要求。象棋发明者

非常聪明，但又很贫穷，他知道这位皇帝很看重其他人的谦逊品质，但却从不这样要求自己。于是他说："尊敬的陛下，我是一个卑微的人，只希望得到小小的补偿来养活我的家人。如果你能仁慈地同意我的请求，那么请在棋盘的第一个方格里赏赐我一粒米，在第二个方格里奖赏我两粒米，在第三个方格里奖赏我4粒米，以此类推，每个方格里的米粒数都是前一个方格的二倍，直到最后一个方格。"象棋发明者的请求看起来十分谦逊，这让皇帝感到吃惊，便吩咐他的仆人去取所需数量的大米。但是，当他们试图计算出到底需要多少大米时，才发现整个国家产出的大米都无法满足象棋发明者的请求。

在这个例子中，我们面对的是一个底数为2的指数函数。棋盘上一共有64个方格，如果每个方格里的米粒数都是前一个方格的二倍，那么总米粒数将是

$$2^0+2^1+2^2+2^3+\cdots+2^{63}=2^{64}-1。$$

它大约是2×10^{19}粒米，重4×10^{14}千克，因为每千克大米有5万粒左右。这是当今世界稻米年产量（约6亿吨或6×10^{11}千克）的500多倍。难怪皇帝不能应允象棋发明者的请求，当然，即使皇帝一开始误以为象棋发明者的要求十分谦逊，我们也不会为此感到惊讶。

第二个例子引自查尔斯·达尔文的开创性著作《物种起源》。在下面这段引自该书的文字中，达尔文提出，动物繁殖固有的指数增长模式必定会受到选择压力的控制，否则仅一个物种的后代就能占领整个地球：

> 自然界中每种生物的数量无一例外地都在快速增加，如果不遭到破坏，一对生物的后代很快就会遍布地球。……大象被视为所有已知动物中繁殖速度最慢的动物。我努力估算了其可能的最低自然增长率：我们保守地假定它在30岁时开始繁殖，一直持续到90岁，其间共生产三对小象。如果是这样，那么在500年后，第一对大象的子孙后代就会有1 500万头存活于世。

尽管达尔文弄错了数字——开尔文勋爵很快就指出了这个错误——但他的想法是正确的。如果改正数字错误，那么我们可以很容易地证明500年后仅有大约1.3万头大象存活于世。如果30年为一代，那么第n代大象的存活数量可以利用$a_n=2\times a_{n-1}-a_{n-3}$计算出来。它意味着每一代存活的大象数量与上一代的比率会迅速收敛至黄金比率——1.618。

指数增长的惊人性质最终证明达尔文的观点是正确的。事实上，尽管100年后存活的大象只有14头，但1 000年后存活着的大象会超过3 000万头，10 000年后（按照进化的标准，10 000年不过是一瞬间）则是1.5×10^{70}头。这个数字足以与宇宙中的粒子总数相提并论，据估计（尽管存在很大的不确定性）宇宙中共有10^{72}~10^{87}个粒子。在这个例子中，指数函数a^n的底数a约等于1.618（n表示世代数）或1.016（n表示年数）。尽管如此，一段很长的时间过后，除了早有心理准备的读者之外，这种增长会让所有人都大吃一惊。

另一个关于指数行为的奇怪性质的例子有些许不同，它来自豪尔赫·路易斯·博尔赫斯的短篇小说《巴别图书馆》（*The Library of Babel*）。博尔赫斯设想了一座图书馆，里面的藏书多到无法想象。每本书有410页，每页有40行，每行有80个字母。这座图书馆藏有所有这种规格且可以用特定字母表中的25个字母书写的书，任何人能想象出的这种规格的书都可以在这座图书馆里找到。那么，这里一定有一本书是详细介绍你的生平的，从你出生的那一刻起到你去世的那一天；而且，这本书有无数个译本，被翻译成现存的所有语言，以及无数种并不存在但十分连贯的语言。有一本书记录了从你出生到你生命中的某一天（由于篇幅限制，该书只能记录到这一天）说的一切话语；第二本书继续记录你之后说的话语，依此类推，直到有一本书记录下你的临终遗言。令人遗憾的是，这些书很难找到，不仅因为大多数书都是胡言乱语，还因为你无法区分这些书，它们实在太相似了。

巴别图书馆的概念令人难以置信，人类有可能写出来的所有（特定规格的）图书竟然都被收藏其中，但是，当你意识到这样的图书馆根本不可能存在，也永远不可能建造出来时，你就不会备感困惑了。它涉及的图书数量如此巨大，以至于任何事物都无法与之相媲美。以立方毫米为测量尺度，可观测宇宙的大小（约为10^{90}立方毫米）甚至比不上80个字符排列组合成的不同行数（约为10^{112}行），更不用说巴别图书馆的藏书数量（$10^{1\ 834\ 097}$）了。

然而，即使是巴别图书馆的令人吃惊的藏书数量，在生物的海量DNA编码组合面前也会黯然失色。我们现在知道，人类和所有其他已知生物的特征，都是通过每个细胞的DNA编码一代代传递下来的。人类基因组包含23对染色体，总共约有60亿个DNA碱基（女性有2×3 036 303 846个，男性的碱基数量略少，因为Y染色体比X染色体小得多）。值得注意的是，人类基因组中99%以上的DNA碱基在每个人类个体身上都完全相同，而物种的所有变异性都是由其余的碱基编码造成的。人类个体基因组之间的差异大多表现为SNP（单核苷酸多态性），指在DNA的某些点位上，不同人的特定DNA碱基通常有不同的值。由于每个染色体对都包含两条染色体（X和Y染色体除外），所以一共有三种不同的可能组合。例如，在一个碱基T和碱基G同时出现的SNP中，有些人可能是T－T碱基对，有些人是T－G碱基对，还有些人是G－G碱基对。含有两个以上碱基的SNP目前比较稀少，在我们的粗略分析中可以忽略。

尽管人类基因组中SNP的确切数量尚不清楚，但据估计有几千万个。在第一批人类基因组测序项目中，其中一个识别出大约200万个SNP。假设这个数字是保守估计，并且所有的基因组变异都是由SNP引起的，我们就可以估算出可能的SNP组合的总数，进而估算出可能有多少不同的人类基因组。我们可以忽略每个人体内的每个染色体对都有两条染色体带来的小麻烦，而只考虑每个人的每个SNP都有三个可能的值。于是，我们知道人类基因组内的SNP一共有$3^{2\ 000\ 000}$种可能的排

列。这意味着，仅因为SNP的变异，就会存在形形色色的人，其数量之大，即使是巴别图书馆的藏书数量也会相形见绌。

当然，如果我们不只是考虑人体内的SNP数量，而是考虑上述大小的基因组中的碱基构成的所有可能的遗传组合，其数量就会多得多。想一想，30亿个DNA碱基可以构成$4^{3\ 000\ 000\ 000}$种可能的组合；再想象一下，在如浩瀚宇宙般的非有效基因组中，有的基因组可以编码出各种各样的神奇生物——这些生物从不存在，也未被创造出来，甚至无法想象。在与现存生物对应不上的无数碱基组合中，有大量的遗传编码可以创造出超乎我们想象的东西，比如飞人、独角兽、超智能的蛇形生物等。

为了简单起见，我们在这里只考虑每个基因组拷贝中4个碱基的可能组合，这是因为这个基因组的空间范围非常大，也没有任何明显的含义，以至于所有的精确计算都毫无意义。特别是像人类这样的二倍体生物，每个染色体对都有两条染色体，因此我们可以认为这个空间的大小接近$4^{6\ 000\ 000\ 000}$。各种可能性多到令人生畏的程度，我们没有也永远找不到探索这些可能性所需的工具，因此我们无法知道这个设计空间内的情况。幸运的话，我们可以在合成生物学领域找到成熟的工具（现在这些工具尚在开发阶段）后，从这片无比深邃的大洋表面掠过。丹尼尔·丹尼特把这个想象出来的设计空间称为“孟德尔图书馆”，这座图书馆的规模非常巨大，巴别图书馆与之相比显得微不足道。

我希望你们现在已经确信指数函数——即使是那些一开始增长非常慢的函数——很快就会达到一个与我们直觉相悖的值。事实证明，我们可以用一条指数曲线准确地描述几乎所有技术领域的增长。在某些特定领域，这种指数增长已得到充分证明。1946年，理查德·巴克敏斯特·富勒发布了一个名为“从绝对宇宙的基本成分——92种元素的发现时序率看工业革命的概况”的图表，清楚地展示了这个过程的指数性质。雷·库兹韦尔在1999年出版的《灵魂机器的时代》（*The*

Age of Spiritual Machines）一书中提出了“加速回报定律”，它指出，各种进化系统（包括许多技术）的变化速度往往呈指数增长。知名度更大的摩尔定律对这种指数进程做了更加量化的说明，它指出，硅芯片上可以安装的晶体管数量会周期性地加倍（我将在下一章详细讨论这个现象）。

指数增长是很多技术发展的特征，但期望这种增长在整个人类历史上保持不变是不合理的。重大的技术变革，比如农业革命和工业革命，可能已经改变了指数的底数，使增长速度在绝大多数情况下都朝着更快的方向发展。

指数增长作为许多技术的一个特征，会导致我们系统性地低估近期和中期的技术水平。我们将在后面的章节中看到，许多技术正在以指数速度发展。在未来几十年里，这种指数速度将深远地改变世界，以至于我们真的无法预测我们的孙辈到我们目前的年纪时的生活方式。

数字技术在即将发生的变化中会起到核心作用。织布机、拖拉机、发动机和机器人改变了田间和工厂的作业方式，数字技术将继续改变我们日常生活中几乎所有任务的执行方式。数字技术已经使打字机和电报成为历史，随着时间的推移，数字技术还将使许多其他技术变得可有可无，甚至被淘汰。就像打字员、银行柜员和报童一样，许多职业和工作岗位也将变得不再有必要，或者数量越来越少。

本书试图对这些发展做出有根据的推测：它们将如何发展，以及如何影响我们的社会、经济和人类。

出生在计算机时代

我出生在计算机时代，是设计、建造、编程、使用和理解计算机的第一代人。按理说，我应该可以理解日益加快的技术变革速度带来的潮流和趋势，但我有时也无法跟上它们的步伐。尽管存在这一局限性，我仍想尽力展望技术发展的可能前景，特别是计算机和生物医学技术的融合发展。

显而易见的是，计算机或者在它之后出现的东西，不仅会取代所有现存的信息传递工具，包括报纸、电视、广播和电话，还将深刻地改变我们解决交通、食品、服装和住房等基本需求的方式。随着我们充分地理解生物的运行原理，我们将有能力用人工或合成的替代品取代越来越复杂的生物系统。我们将拥有改变环境并最终改变我们身体的能力，这些能力又将深刻地影响我们的生活方式，我们已经看到了年轻人的生活和互动方式正在发生这种深刻变化的迹象。我们可能会怀念过去，那时人们面对面聊天而不是发短信，去公园散步而不是浏览网页，但是，不管你喜不喜欢，这些变化都只是一个序幕。情况可能并不像看上去的那么糟糕。与活在今天的任何人相比，22世纪的每个普通人可能会拥有更多的自由和选择，也更有能力去创造新东西，他们将获得我们做梦也想不到的知识和技术。然而，就像今天的生活方式会让19世纪早期的人感到陌生一样，22世纪的日常生活也会让我们感到陌生。

本书试图带你一窥接下来会发生的事情。我无法预测100年后技术会带来什么，我认为其他人也做不到。但我可以尝试根据现有的技术做出推断，并预测100年后可能会出现的某些事物。

为便于读者理解，我不会一下子就跳到100年后。我将从介绍若干关键领域的近代和更早的技术发展史、趋势及现状入手，推断它们近期的发展状况。我们可以合理地预测在不久的将来可以实现哪些技术发展，这会给予我们足够的勇气去猜测日益加速的技术发展将带来什么样的奇迹。未来肯定和我们想象的不一样，甚至会比我们今天的预测更让人觉得陌生。相较过于大胆，我们更有可能犯过于保守的错误。

我打算从引发当下的数字革命的技术开始我们的讨论。数字革命始于电的发现，并随着晶体管的发明而加速。晶体管看似是一个小物件，但它可能是人类有史以来发明的最具革命性的一项技术。

第3章

从麦克斯韦到互联网

多项重要技术都在19—20世纪得到了迅速发展，包括电、电子和计算机，也包括生物技术、纳米技术及其衍生技术。事实上，正是这些技术的指数发展速度，推动了近几十年来的科学和经济发展。这些技术和其他技术还有可能实现更快的发展。

改变世界的4个方程

在17世纪之前，与电有关的现象（例如，磁性、闪电和会放电的鱼），一直是科学奇闻。18世纪，本杰明·富兰克林首次对电进行了广泛和系统的研究，尽管在他之前也有其他科学家研究过电现象。据说，在1752年进行的一项实验中，富兰克林利用风筝从暴风雨中捕捉到了电能。

亚历山德罗·伏打于1800年发明了电池，之后很多实验都可以利用电流来完成。汉斯·克里斯蒂安·奥斯特和安德烈·玛丽·安培发现了电流和磁场之间关系的多个性质，但真正使人们理解电并把电变成一种有用技术的是迈克尔·法拉第的研究成果。法拉第对电流周围的磁场进行了广泛研究，为电磁场的概念奠定了基础。他熟悉电阻、电容、电感器等基本电路元件，而且研究了电路的工作原理。他还发明了第一台电动机和发电机。电气技术在19世纪得到实际应用，主要归功于他的努力。法拉第想象在电气装置和磁性装置周围存在某种场（他称之为电紧张态），电磁现象是由这种场的内部变化引起的。

不过，法拉第几乎没有受过正规的数学训练，因此，建立方程并用它们来控制电磁场行为的任务就落到了詹姆斯·克拉克·麦克斯韦身上，现在这些方程就是以他的名字命名的。1864—1865年，麦克斯韦（获悉了法拉第的研究成果）发表了他的研究结果，指出电场和磁场是电和磁性的基础，而且它们都可以在空间中以波的形式传播。此外，他认为光本身就是一种电磁波，因为它与电磁场的传播速度相同。这些成果的数学公式形成了一组方程，它们是科学史上最著名和最具影响力的方程之一。

麦克斯韦最初在《电磁通论》中提出这个方程组时，它一共包含了20个不同的方程。（麦克斯韦没有现代数学工具，而我们可以借助这些工具，用简单的形式描述复杂的数学运算。）当时，只有少数几位物理学家和工程师能理解麦克斯韦方程组的全部含义，因此它们受到许多科学界人士的极大怀疑，其中就包括威廉·汤姆森（后来的开尔文勋爵）。许多物理学家都积极投身到研究和发展麦克斯韦方程组的工作中来，历史学家布鲁斯·亨特在一本同名著作中称他们为“麦克斯韦学派”。

很快，奥利弗·赫维赛就理解了麦克斯韦方程组的潜在重要性。在他一生中的大部分时间里，赫维赛都致力于重构麦克斯韦方程组，最终使其变成现在物理学家和工程师熟悉的4个方程。我们可以在T恤和汽车保险杠上看到它们，尽管大多数人都识别不出它们或者忘记了它们的含义。在加州大学伯克利分校的电报大道上、麻省理工学院里和其他许多地方，我们都可以买到印有图3－1所示图案的T恤。这些T恤不禁让人想起路德维希·玻尔兹曼的问题：“这些符号难道是上帝亲手写下的吗？”这句话引自歌德的《浮士德》，“这些符号”指的是麦克斯韦方程组。

在麦克斯韦方程组的优美而又令人费解的数学形式背后，隐藏着控制电磁场行为的定律。我们今天使用的几乎每一种电气或电子设备中，都存在电磁场。麦克斯韦方程组描述的是电场（E）和磁场（B）的相互作用方式，以及它们与包括电荷密度（ρ）和电流密度（J）在内的其他物理实体之间的关系。参数ε_0与μ_0是物理常数，分别叫作自由空间的电容率和磁导率，它们的关系方程是$\mu_0\varepsilon_0=1/c^2$，其中c是光速。

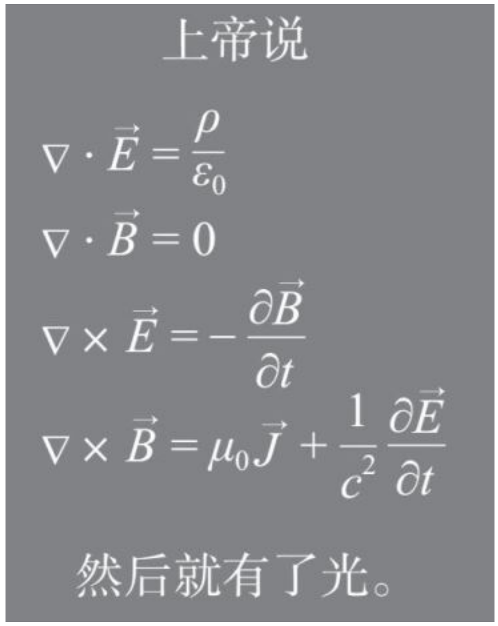


图3－1 根据奥利弗·赫维赛重构的麦克斯韦方程组与一件流行的T恤，我们可以解开世界起源的秘密

场是一个数学构造，它为空间中的每个点都赋予了特定值。对于标量场，这个值是一个数字；对于向量场，这个值是一个有方向和强度的向量。电场、磁场和引力场都是向量场，空间中的每个点都是由它们的方向和振幅定义的。麦克斯韦方程组之所以很难被同时代的人理解，其中一个原因是，他们很难想象或理解场，甚至更难理解场波在真空中的传播方式。

符号“∇”是一个数学运算符，它有两个不同的含义。如果它和场之间的运算符是“·”，那么它表示场的散度。散度指的是，在以某一点为中心的任意小的体积内，向量场的外向通量的体积密度。如果一个点的散度是正值，就说明这里有场的外向通量。引力场的散度是质量，电场的散度是电荷。电荷可以是正的，也可以是负的，但没有负的质量（至少我们还未发现）。如果某一点有正电荷，就会创造出一个出射场；如果有负电荷，就会创造出入射场。

如果∇和场之间的运算符是“×”，那么它表示场的旋度，也就是场在特定点的旋转方式示值。假设场内有一个很小的球，用来表示水等流体的运动情况，那么场的旋度可被视为描述小球旋转运动性质

的向量，这是因为流体流经各个面的速度略有不同。旋度由一个与小球的旋转轴同向的向量来表示，其长度与转速成正比。

麦克斯韦方程组中的第一个方程可以表述为，从某个体积离开的电场总量等于该体积内的电荷总量。这是对电场由带电粒子产生这个事实的数学表述。

第二个方程指出，从某个体积离开的磁场总量必定等于进入该体积的磁场总量。在这个方面，电场和磁场不同，因为电场是由带电粒子产生的，但并不存在可以产生磁场的磁单极子。因此，磁场总是闭合回路，进入某个体积的磁场总量必定等于离开该体积的磁场总量。

第三个方程指出，一个闭合回路周围积累的电势差，即电压差，等于通过该回路封闭区域的磁通量随时间发生的变化。

第四个方程指出，流经某个面积的电流及电场变化与该面积周围分布的磁场成正比。

随着时间的推移，人们对电场和磁场之间关系的理解，促使电发展成20世纪最有用的技术。1888年，在德国卡尔斯鲁厄大学的前身任教的海因里希·赫兹率先通过实验证明了麦克斯韦预言的电磁波的存在，以及电磁波可用于远距离传输信息。

现在，我们依靠电动机和发电机驱动电器，生产我们购买的消费品，加工和保存我们食用的大部分食物。电视、电话和计算机依靠电场和磁场发送、接收和存储信息，许多医学影像技术也建立在麦克斯韦方程组的基础之上。

电气工程师和物理学家每天都会使用麦克斯韦方程组，尽管有时是通过不同或者简化的形式。举一个简单的例子，我们就可以生动地理解麦克斯韦方程组是如何定义电路行为的。我们来看第三个方程，它表明当穿过闭合电路封闭区域的磁场不变时，在该电路周围测得的电压和为零。图3－2所示电路中包括一个作为电压源的电池、一个电容器和一个电阻，应用第三个方程可得

$$V_R + V_K - V_C = 0。$$

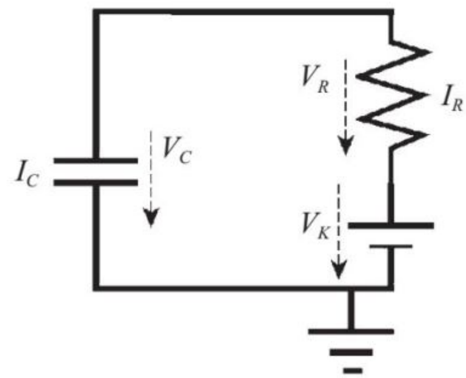


图3－2 由一个电容器、一个电阻和一个电压源组成的电路

这个表达式是通过计算电路（电流沿顺时针方向流动）周围的电压降之和得出的。电压降V_C需要添加一个负号，因为它与电流方向相反；电压V_K（电池电压）和V_R（电阻电压）与电流方向相同。结果表明，通过电容器I_C的电流与通过电阻I_R的电流大小相等（方向相反），因为这个电路中所有的电流都会流经导线。

麦克斯韦的第四个方程表明，通过电容器的电流与电容器极板间的电场随时间的变化率成正比，由此得到表达式$I_C = C\dot{V}_C$，这是因为电场（在本例中是$\vec{E}$）与电容器极板间的电压差V_C成正比。在这个方程中，V_C上方加点表示这个量随时间的变化率。在数学上，它被称为关于时间的导数，也可以表示为dV_C /dt。

1827年，另一位德国物理学家乔治·欧姆描述了电阻上的电流I_R和电压V_R之间的线性关系。欧姆定律指出，根据电阻R的值，这两个变量之间存在线性关系$V_R=RI_R$。把这些表达式放在一起可以得到

$$-C\dot{V}_C = (V_C - V_K)/R。$$

这个表达式是应用麦克斯韦方程组和欧姆定律得到的直接结果，也是一个含单一未知数V_C（电容器上的电压，会随时间发生变化）的

微分方程。表达式中的所有其他参数都是已知的固定物理量。这类微分方程在变量及其随时间的变化率之间建立了联系。上面这个微分方程可以用解析法或数值法求解，从而得出电容器上的电压随时间的变化率。通过类似的分析，我们可以推导出包含几千个甚至几百万个电子元件的复杂电路的行为。

我们现在知道，大脑中的神经元正是利用麦克斯韦方程组描述的电磁场来表演它们的“魔法”的。目前，速度飞快的计算机通过求解麦克斯韦方程组的方式，对人脑和电路进行模拟。

物理学的世纪

20世纪被称为物理学的世纪。虽然电磁学知识在19世纪得到了迅速发展，而且在1900年以前人们就对引力形成了重要见解，但物理学是塑造20世纪的重大技术进步的驱动力。在这100年里，人们将强核力、弱核力、电磁力和引力放到一起，构建出我们现在看到的支配宇宙的4种相互作用的完整体系，第一次全面认识了物质和能量之间的关系。

阿尔伯特·爱因斯坦于1905年发表的“奇迹年论文”为20世纪开了一个好头。爱因斯坦在狭义相对论、物质和能量当量、光电效应及布朗运动等方面做出的开创性贡献，改变了物理学。这些出版物加上我们对电磁学日益增长的理解，推动了一系列的技术发展，并带来了今天的计算机和通信技术。

20世纪20年代诞生了量子力学，这种关于光和物质的反直觉理论是许多位物理学家的研究成果。路易·维克多·德布罗意是在量子力学领域做出贡献的第一批物理学家之一，他断言粒子的行为像波，电磁波的行为有时也像粒子。欧文·薛定谔（为量子力学奠定概率基础的第一人）和沃纳·海森堡（提出不可能同时精确地测量粒子的位置和动量，即不确定性原理）也做出了其他重要贡献。尽管量子力学已被证明是有史以来最坚实和经受了最精确检验的物理学理论之一，但这些革命性理论提出的哲学问题至今仍未得到解答。

第二次世界大战后，朱利安·施温格、理查德·费曼和朝永振一郎各自运用已有的量子理论提出了解决数值难题的方法，为建立强有力的量子电动力学理论开辟了道路。费曼还是一位才华横溢的故事讲

述者，也是有史以来最具影响力的科学作家之一。他的《QED：光和物质的奇妙理论》深入浅出、通俗易懂，少有书能做到这一点。

高能物理学带来了一系列新发现，一大批亚原子粒子丰富了粒子物理学的宇宙。基本力通过粒子来介导的观点（光子介导的是电磁力，介子介导的是强核力和弱核力）在实验中得到了验证。组成物质和能量的粒子除了众所周知的电子、中子、质子和光子以外，还有许多其他的奇异粒子，包括正电子（电子的反物质版本）、反质子、反中子、π介子、K介子、μ子、τ子和中微子等。最初这些粒子只是一种假设，或者主要是通过宇宙线留下的电离尾迹被发现的，但随着粒子加速器（比如欧洲核子研究组织和费米实验室的加速器）的出现，越来越多的粒子被制造出来。甚至一些更奇异的粒子，包括弱相互作用大质量粒子（WIMP）、W粒子、Z^0粒子和难以捕捉的希格斯玻色子，也加入了这个队伍。这使得在未来的许多年里，理论物理学家将忙于探索物理学的统一理论。

若干奇异的理论被提出来，试图用尽可能少的自由参数把引力与其他三种力（强相互作用、电磁相互作用和弱相互作用）统一起来。这些理论尝试取代标准模型，后者统一了电磁相互作用、强相互作用和弱相互作用，并涵盖了量子电动力学和量子色动力学。其中最流行的是弦论，该理论认为，基本粒子可被看作在某个高维空间中振动的弦。例如，流行的M理论要求十一维时空，而且这个要求几乎不可能降低了。到目前为止，所有理论在预测观察到的现象时，除非事后小心地调整参数，否则都无法取得成功。

在我写作本书期间，位于瑞士的欧洲核子研究组织的大型强子对撞机，代表了我们理解高能物理学世界的最新努力。2012年，欧洲核子研究组织的科学家探测到了难以捕捉的希格斯玻色子，在此之前标准模型预测的所有粒子中只有它未被观测到。由于这种粒子与光子（介导电磁力且无质量）及大质量的W玻色子和Z玻色子（介导弱核

力）的不同特性有关，再加上之前一直未被观测到，人们对它就更加好奇了。

现有的物理学理论非常不直观，即使对专家来说也是如此。量子物理学的一个基本信条是，我们无法预测观测的真实结果，但可以计算出它的概率。这个简单的事实产生了若干非常反直觉的结果，比如著名的薛定谔的猫悖论和爱因斯坦–波多尔斯基–罗森佯谬。

薛定谔提出了一个思想实验：一只猫被放在一个密封的盒子里，它是生是死取决于一个放射性原子的状态，这个原子可以通过某种装置控制有毒物质的释放。薛定谔指出，在外部观察者进行观测之前，这只猫既是活的又是死的，即处于与一个随机的亚原子事件有关的宏观量子叠加态，而该随机事件的发生概率取决于量子力学定律。

薛定谔设计这个思想实验的目的是，说明量子力学的标准诠释产生的悖论，在某种程度上它也是对爱因斯坦–波多尔斯基–罗森佯谬做出的一个回应。爱因斯坦–波多尔斯基–罗森佯谬的三位作者设想，如果两个粒子处于量子纠缠态，测量一个粒子的性质，就会立刻使另一个粒子的相关性质取某个特定值。这个概念性问题之所以会出现，是因为许多物理学实验都会创造出成对的纠缠态粒子，它们在被创造出来之后会彼此远离。测量其中一个粒子的状态，就会迫使另一个粒子的状态取某个特定值，即使两者之间有几光年的距离。例如，如果一个自旋为零的粒子，比如光子，衰变成一个电子和一个正电子，由于自旋守恒，这两个粒子都必定有自旋，而且自旋方向相反。但是，只在测量其中一个粒子时，它的自旋才是已知的。在那一刻，另一个粒子的自旋将取相反的值，无论它们的空间距离有多远。根据相对论，这种“鬼魅般的”瞬间超距作用被视为不可能，因为它可被用于以超光速的速度传输信息。后验诠释表明，事实上纠缠机制并不能用于传输信息，但爱因斯坦所谓的“鬼魅般的超距作用”机制仍然令人费解。

尽管上述悖论和量子力学带来的其他悖论引发了激烈的讨论，但如果不是因为量子力学对现实世界产生的影响，它们可能不会受到公众的关注。如果物理学的发展没有达到现在的程度，我们日常生活中使用的许多东西都不会发生什么变化。我们仍然会像我们的祖先那样饲养家畜，种植作物，乘坐汽车、火车和飞机去往各地。这些交通工具从某种程度上说是第一次工业革命的产物，即使没有现代物理学，它们也会被开发出来。

但是，有两个事物属于明显的例外。第一个例外在1945年8月6日引起了全世界的关注，那一天第一颗原子弹在广岛上空爆炸，其威力相当于12 000~15 000吨TNT（三硝基甲苯）炸药。若干年前，由物理学家J. 罗伯特·奥本海默领导的曼哈顿计划接受了一项使命，探索是否可以利用爱因斯坦的$E=mc^2$方程制造出威力远胜从前的炸弹。众所周知，基于裂变和聚变的炸弹被开发出来。

第二个例外是一种叫作晶体管的十分简单的设备。一开始，它并未引起人们的关注，但40年后，它在人类社会中掀起了前所未有的深远革命。

晶体管、芯片和微处理器

晶体管类装置的第一个专利可以追溯到1925年，人们仅通过实验研究就发现了使晶体管成为可能的效应。但是，对量子力学以及由此产生的固体物理学领域的理解使我们意识到，利用半导体的电学性质可以实现借助电阻、电容器等简单的电气元件无法实现的行为。

1947年，贝尔电话实验室的研究人员观察到，当电触头被放置在锗晶体（半导体）上时，输出信号的功率比输入信号的功率大。因为这个堪称有史以来最重要的发现，威廉·肖克利、约翰·巴丁和沃尔特·布拉顿获得了1956年的诺贝尔物理学奖。肖克利预见到这种装置会有许多重要的应用，于是在加州山景城建立了肖克利半导体实验室。晶体管引发的这场巨大变革最终推动了现代计算机技术的出现，肖克利可以说是“第一个吃螃蟹的人”。

晶体管是一种简单的装置，有三个端子，其中一个控制输入。通过改变控制输入端子的电压，可以让通过晶体管的另外两个端子的电流发生很大的变化。这可被用来放大麦克风捕捉到的声音，或者产生强大的无线电波，还可以被用作控制开关。

第一类晶体管是双极结晶体管，它的控制输入端子（被称作“基极”）中也会有较小的电流。最终，其他类型的晶体管主宰了这项技术。金属－氧化物－半导体场效应晶体管（MOSFET）依据的是不同的物理学原理，但基本效果相同：改变控制输入端子（在这种晶体管中被称作“栅极”）的电压，流经晶体管的另外两个端子（在这种晶体管中分别被称作“源极”和“漏极”）的电流就会发生显著变化。仅仅依据这个简单的描述，你可能会很难理解为什么这样一个装置能使人类社会在过去30年里发生如此巨大的变化，在未来的几十年里还会

引发更大的变化。20世纪初发明的真空管表现出与晶体管相似的行为，实际上两者也有许多相同的应用。即使在今天，一些音频放大器和利基应用仍在使用真空管。虽然晶体管比真空管更可靠也更不容易出故障，但两者之间的关键区别在于，晶体管可以做到尺寸非常小，大量使用，而且单位成本非常低。人类花了几十年的时间，才将晶体管的这些优势充分挖掘出来。英国工程师杰弗里·杜默是第一个提出将诸多晶体管封装成集成电路想法的人，把这一想法变成现实的人则是杰克·基尔比和罗伯特·诺伊斯，而且晶体管之间不需要靠导线来连接。

被紧密封装为集成电路的晶体管有许多用途。它们可被用来放大、操控和产生模拟信号，许多设备都利用晶体管来实现这一目的，比如无线电和电视接收器、扩音器、移动电话和GPS（全球定位系统）接收器。电路设计人员能够把晶体管封装在体积很小的放大器和其他信号处理元件中，而真空管则做不到这一点，不仅会占据大量空间，而且会重达好几吨。个人移动通信技术的发展在很大程度上正是得益于晶体管的这种特殊应用。

不过，晶体管技术之所以能对我们的生活产生巨大影响，更多是因为包含大量晶体管的集成电路易于进行大规模生产，而且晶体管可以充当控制开关去处理数字信息。数字计算机处理的数据是二进制形式。数字、文本、图像、声音和所有其他类型的信息，都以一长串比特（由二进制数字0和1构成）的形式存储。这些二进制数由数字电路处理，并被存储在数字存储器中。数字电路和存储器是由逻辑门构成的，逻辑门又是由晶体管构成的。例如，一个“与非门”有两个输入和一个输出。当且仅当两个输入都为1时，它的输出为0。与非门可能是最简单的门电路之一，由4个晶体管构成，如图3－3（a）所示。图中有4个金属－氧化物－半导体场效应晶体管：底部两个为N型，顶部两个为P型。N型金属－氧化物－半导体场效应晶体管在栅极处于高压时就像一个闭合的开关，在栅极处于低压时则像一个打开的开关。P型金属－氧化物－半导体场效应晶体管（可根据栅极上是否有一个小圆

圈加以区分）正好相反：当栅极处于低压时，它就像一个闭合的开关；当栅极处于高压时，它则像一个打开的开关。

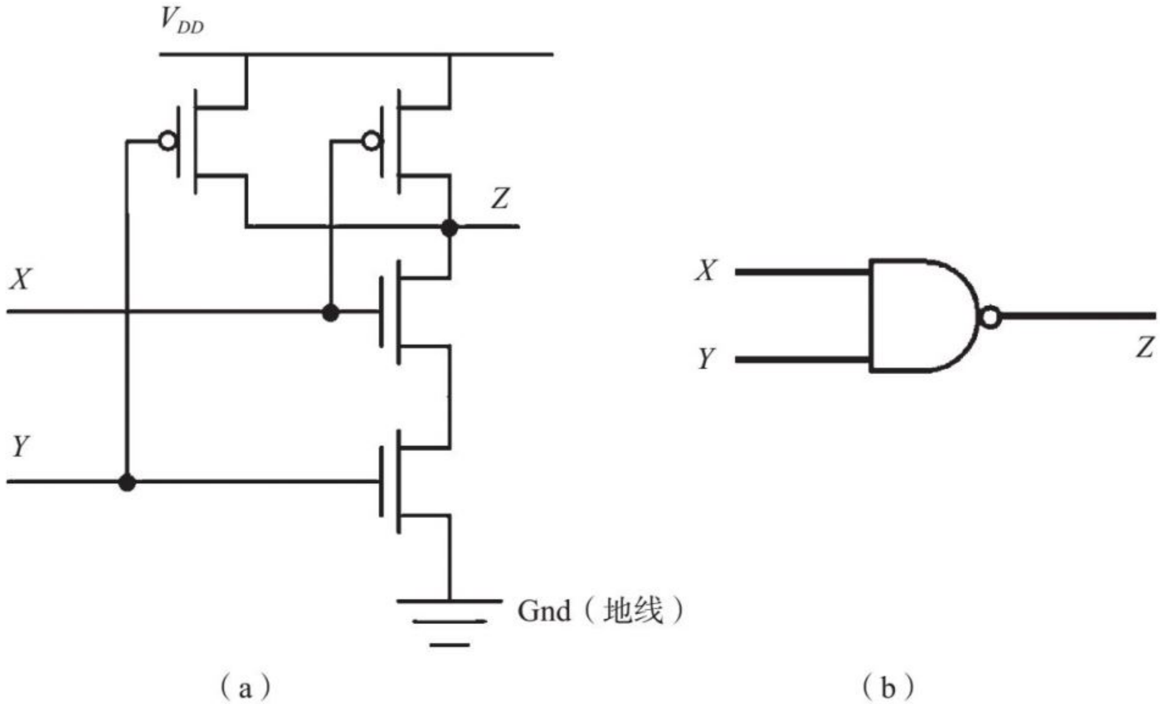


图3－3 （a）由金属－ 氧化物－ 半导体场效应晶体管构成的与非门；（b）与非门的逻辑符号

如果X和Y（晶体管的控制输入）的逻辑值均为1（通常为供电电压V_{DD}），那么图3－3（a）底部所示的两个晶体管起到闭合开关的作用，将输出Z连接到地面，逻辑值为0。如果X或Y（或X和Y）的逻辑值为0（通常为地面，或电压为0伏特），那么图3－3（a）底部所示的两个晶体管中至少有一个起到打开开关的作用，而上方的两个晶体管（或其中之一）则会起到关闭开关的作用，使Z的值上升到V_{DD}（对应的逻辑值为1）。这相当于计算$Z=\overline{X\wedge Y}$，其中的横线表示否定，∧表示逻辑操作“与”。这个函数的真值表可参见表3－1。

表3–1 逻辑函数与非门的真值表

X	Y	Z
0	0	1
0	1	1
1	0	1
1	1	0

除了与非门之外，还有许多其他类型的逻辑门可用于计算不同的逻辑函数。反相器输出与其输入相反的逻辑值，实际上就是一个简化的与非门，它只包含两个晶体管，一个是P型，另一个是N型。与门计算输入逻辑值的“与”，只当两个输入都是1时才会输出1。在与非门后面加一个反相器，就可以构成与门。其他类型的逻辑门，包括或门、异或门和或非门，都可以通过晶体管和基本门的不同排列来构建。更复杂的数字电路就是由这些简单的逻辑门构成的。

与非门在某种意义上具有特殊性，只需要使用与非门，就可以构建出任何逻辑函数。事实上，把与非门结合起来，可以计算任何逻辑函数或二进制函数。这种特性源自与非门的完备性，也就是说，任何逻辑函数，无论多么复杂，都可以用与非门来创建。例如，与非门可用于建立两位异或函数，当只有一个输入位为1时，该函数的值为1。与非门也可用于建立三位多数函数，当两个或两个以上的输入位为1时，该函数的值为1。图3－4展示了如何利用与非门建立两位异或函数（当只有一个输入为1时，函数的值为1）和三位多数函数（当至少有两个输入为1时，函数的值为1）。

事实上，完全由与非门构建的电路可以计算二进制数的加减乘除，以及逻辑电路可以计算的任何其他函数。在实际应用中，几乎所有复杂的数字电路都是由与非门、或非门和反相器构建的，因为这些门不仅可以计算逻辑函数，还可以再生电信号的电平，因此可再次被用作其他逻辑门的输入。从概念上讲，一台完整的计算机可以仅由与非门构建而成；事实上，很多计算机就是这样建造的。

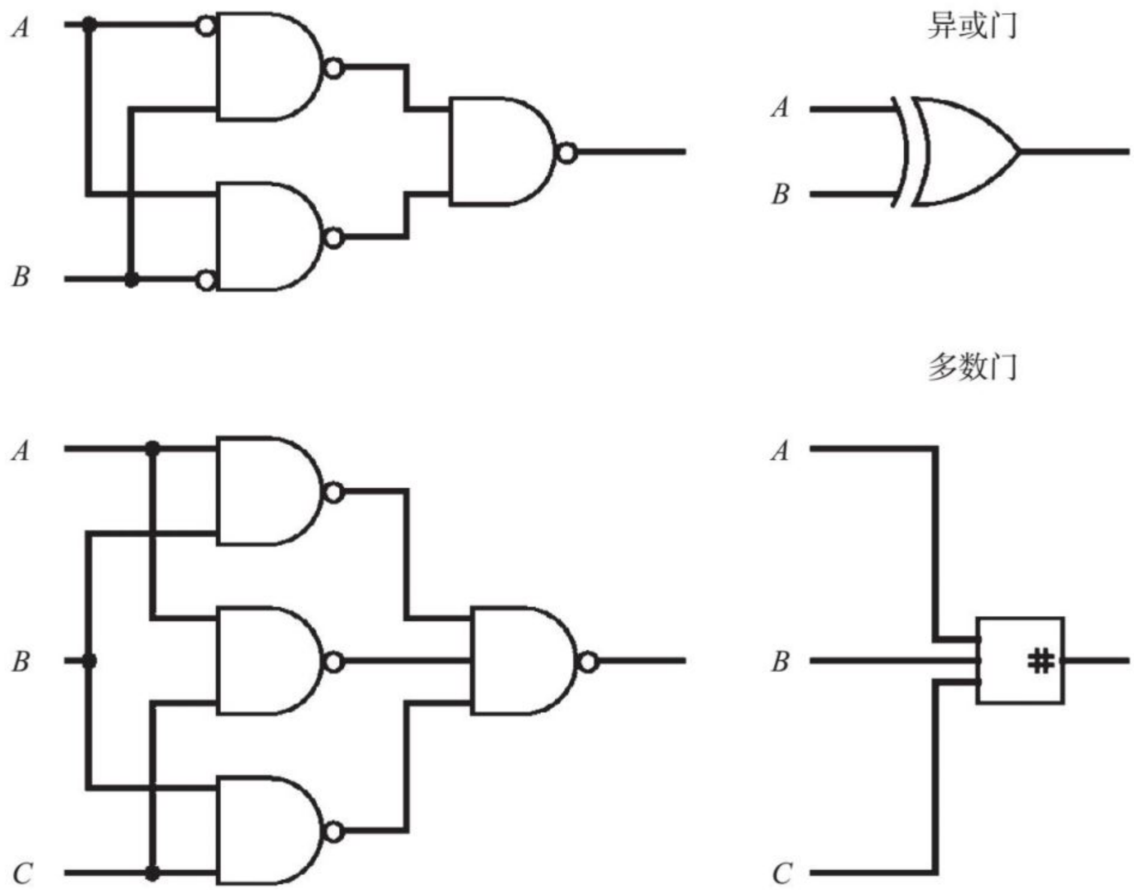


图3－4 由与非门组成的异或门与多数门

计算机内部只能处理二进制数。虽然我们习惯使用由数字0~9构成的十进制数，但以10作为基数并没有什么特别的原因。这可能是因为人类有10根手指，以10为基数似乎是自然而然的事。以10为基数的数字其实就是10的幂次方的加权和的简洁表达，例如，121可表示成

$$1\times10^2+2\times10^1+1\times10^0,$$

因为它的每一位均对应10的某次幂。

如果基数不是10，那么只需要将上式中的10替换成相应的基数即可。如果基数小于10，则不需要10个符号就可以表示所有数字。同样是121这个数，如果基数不是10而是4，它就会变成

$$1\times4^3+3\times4^2+2\times4^1+1\times4^0$$

（也可以写成1321_4，下标表示基数）。当基数为2时，我们要用的就是2的幂次方，而且数字只有两个，即0和1，它们可以方便地表示成两个电压电平（例如，0伏和5伏）。基数为10的121_{10}转换成基数为2的数，就会变成1111001_2，它表示

$$1\times2^6+1\times2^5+1\times2^4+1\times2^3+0\times2^2+0\times2^1+1\times2^0。$$

利用计算预期函数的逻辑电路，可以完成二进制数的算术运算。例如，如果我们使用的都是四位数，并且希望计算7_{10}与2_{10}的和，就必须先将它们转换成基数为2的形式，即01112和00102，再求和。

我们在小学都学过加法算法，如图3－5所示。该算法从最右列开始，依次把每列上的数字相加，并将某一列的进位位写到其左边一列的上方。二进制加法唯一的区别在于，每一列只有4种可能的输入组合，因为每个数字只能是0或1。但是，由于进位位也有两个可能的取值（要么有进位，要么没有），所以每一列共有8种可能的组合。表3－2列出了这8种组合，以及输出C和进位位Cout的预期值。进位位一定要与其左边一列的数字相加，所以它既是某一列（Cout）的输出，又是其左边一列（Cin）的输入。

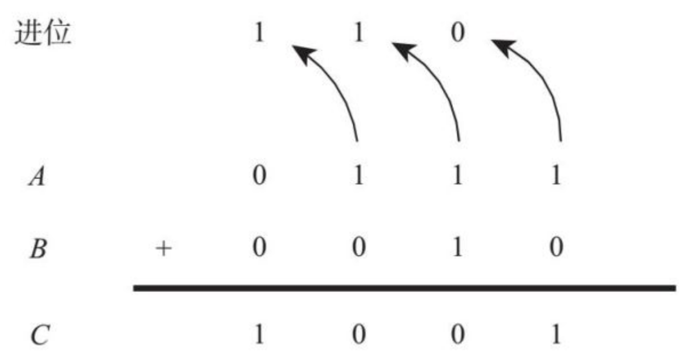


图3－5 二进制数0111和0010相加

表3–2 两个二进制数相加的逻辑函数。C列给出的是结果位的值，Cout列给出的是进位位

A	*B*	*Cin*	*C*	*Cout*
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

仔细检查很容易就会发现，函数C的值是由三个输入位的异或给出的，当有奇数个位为1时，函数值为1。函数Cout的值则是由这三个输入位的多数函数给出的。

因此，图3－6左边的电路计算的是由A、B和Cin这三个输入得到的输出（C）和进位位（Cout）。更有趣的是，将这4个电路连接在一起，就可以得到图3－6右边所示的四位加法器。在这个四位加法器中，最上面的一位加法器将数字A和B的最低有效位相加，进位位通过加法器链传送。该电路利用图3－5所示的算法（和本例中的值）进行两个四位数字的加法运算。

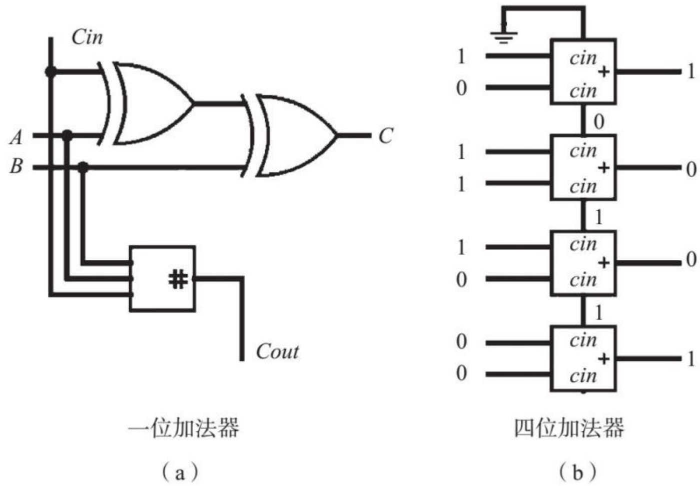


图3－6 （a）一位加法器；（b）对二进制数0111和0010进行加法运算的四位加法器

利用这些基本结构可以构建更复杂的电路（例如，计算两个二进制数的乘积）。乘法可以用专门的电路来完成，也可以用我们小学学过的算法，把相同的数相加多次。加法器、乘法器等基本结构可以相互连接，形成通用数字电路，对二进制数进行各种组合的基本运算。于是，不起眼的晶体管成为计算机行业的“苦力”，帮助人们建造出物美价廉的加法器和乘法器（1651年，托马斯·霍布斯提出一个猜想，即人类所有的推理和记忆活动都建立在加法器和乘法器的基础之上）。

晶体管和逻辑门被安装在电路中，可以长时间存储二进制值，也可以用于构建计算机存储器。我们今天使用的所有计算机，都拥有这种可以存储几十亿甚至几万亿比特的存储器。因此，利用晶体管构建的通用电路不仅可以完成所有可能的逻辑运算，而且速度快、成本低、效率高。一个非常复杂的数字电路可以根据指令，将一个存储位置上的内容与另一个存储位置上的内容相加，并将结果存储在第三个存储位置上。可以足够灵活地执行这些和其他类似操作的数字电路，被称为中央处理器（CPU）。中央处理器是我们今天使用的所有计算机和几乎所有先进电子设备的大脑，它执行的程序不过是一长串非常简单的运算。（我将在下一章解释阿兰·图灵、约翰·冯·诺依曼等人的开创性研究，是如何让中央处理器变成现代计算机的大脑的。）

第一代数字计算机由真空管制成的逻辑门互相连接构成。这些机器体积庞大，速度缓慢，而且不可靠。1946年诞生的第一台电子数字积分计算机ENIAC包含17 000多个真空管，重量超过27吨，占地面积大于600平方英尺(注)。

基于分立式晶体管的计算机成为常态之后，大幅减少了占地面积和耗电量，但真正的突破是英特尔公司的设计师实现的。他们发现一块芯片就可以胜任中央处理器的工作，这种芯片后来被称作微处理器。第一款单片机中央处理器（4004处理器，发布于1971年）可以处理四位二进制数，有2 300个晶体管，重量不足1克。现在的高端微处理器有超过30亿个晶体管，它们都被封装在一个邮票大小的区域里。

现在，晶体管可以每秒约150万亿（1.5×10^{14}）个的速度大规模生产。迄今为止，晶体管的总产量已经超过3×10^{21}个，这个数字完全可以与地球上的沙粒数量相提并论。几年之内，我们制造的晶体管数量就将超过目前存活于世的所有人类大脑的突触数量。

自1971年以来，微处理器中的晶体管数量迅速增长，符合一条被称为摩尔定律的近似指数曲线。1965年，英特尔的联合创始人戈登·摩尔率先发现，当价格不变时集成电路上可容纳的晶体管数量随时间变化呈指数增长，大约每两年就会加倍。图3－7描述了自4004处理器问世以来，英特尔微处理器中晶体管数量的增加情况。注意，为方便起见，晶体管的数量是以对数尺度表示的。尽管该图展现的只是一家供应商的少量微处理器的情况，但它显然符合摩尔定律。在这个案例中，微处理器中的晶体管数量在41年内增加了2^{20}倍，大约相当于每两年增加一倍。

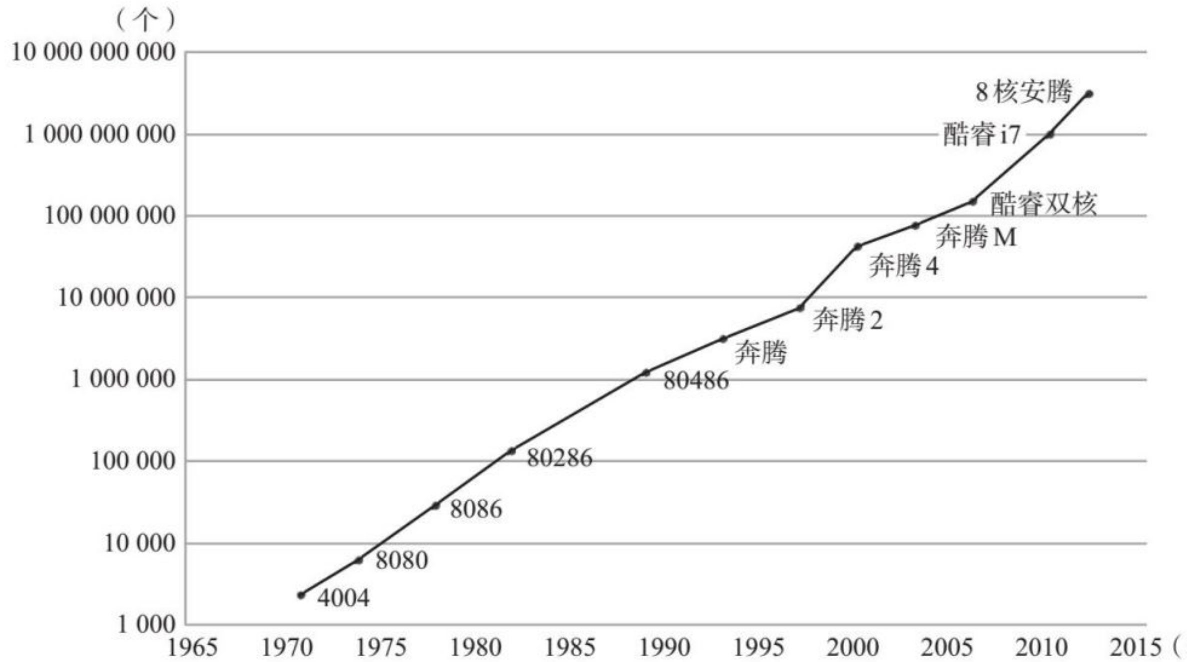


图3－7 英特尔微处理器中的晶体管数量的变化情况

数字电子设备进化的很多量度都遵循与摩尔定律类似的规律。处理速度、存储器容量和传感器灵敏度都在以指数速度提升，而且与摩尔定律的预测非常接近。数字电子产品几乎影响着我们生活的方方面面，根本原因就在于这种指数增长。事实上，摩尔定律和数字技术的指数进化引发的诸多事件，近几十年来已经深刻地改变了人类社会。

其他数字技术也在以指数速度提升，尽管在某种程度上它们不受摩尔定律的影响。克莱德定律指出，磁盘的特定区域内可以存储的比特数大约每13个月就会增加一倍。拉里·罗伯茨详细记录了通信设备的改进数据，他发现在过去的10多年里，固定通信容量的单位成本呈指数下降趋势。

就上述关于摩尔定律的例子而言，由于处理器的速度也在不断加快，因此处理器技术的进步幅度甚至比图3－7展示的更加显著。仅从处理器性能的角度看，随着晶体管数量的增加、处理器速度的加快，计算能力也会不断提升。因此，1971—2012年，处理器的计算能力提高了约300亿倍，相当于每14个月增加一倍。

数字技术的进步带来的这种指数增长，是其他科学领域都无法望其项背的，但有一个领域例外，那就是DNA测序。近几十年来，交通、能源和建筑行业也取得了显著进步，然而，这些行业都没有实现像半导体技术那样的指数增长。我们来看一下汽车燃油效率。在与上述讨论大致相同的时期内，乘用车的燃油效率从每加仑[注]20英里提高到每加仑35英里。如果乘用车的燃油效率在过去40年里经历了与计算机相同的发展速度，那么一加仑汽油可以让一辆普通的乘用车绕地球行驶100多万圈。

集成电路技术的指数速度的进步，推动了信息和通信技术的发展。随着时间的推移，由数字电路技术支撑的高速网络把计算机相互连接在一起，形成了万维网。这个巨大的网络将绝大多数的现存计算机连接起来。

有显著的证据表明，25年后摩尔定律将逐渐失去效力——可以封装到一块芯片上的晶体管数量的增速不再像过去那么快了。但是，到那时其他技术可能会发挥作用，推动计算机的性能继续增长（尽管增速有所减缓）。

互联网的崛起

数字计算机刚出现时，主要应用于科学和军事领域，以取代“人类计算机”。在数字计算机出现之前，科学和军事领域的表格是由大量的人力计算的，这些人被称为人类计算机。在20世纪60年代早期，大型计算机（可占据整个房间）开始应用于商业领域。但是，直到最近几十年人们才清楚地认识到，计算机必将成为有史以来普及程度最高的设备。

在历史上，严重低估计算机未来发展前景的评价比比皆是。一个可能是杜撰的故事称，1943年IBM（国际商业机器公司）的创始人托马斯·沃森指出，全球市场的计算机需求量可能是5台。据说，直到1977年，美国数字设备公司的董事长兼创始人肯·奥尔森还评论道，“认为任何人都想拥有一台家用计算机，这种想法是没有道理的。”1981年，微软公司董事长比尔·盖茨声称，640千字节的存储器对所有人来说应该足够了。这些预测都大大低估了计算机技术的发展及其在现代世界的普及程度。

但是，直到万维网（1989年有人提议建立万维网，将不同计算机中的文档连接起来，以便读者在查阅其中一份文档时可以方便地看到其他相关文档，但这个提议几乎没有引起任何人的关注）出现，计算机才进入了大多数人的日常生活。后来，万维网这个概念的力量被互联网（利用高速通信设备连接而成的巨大计算机网络，遍布全世界）全部释放出来。20世纪70年代初，一个研究团队提出了一系列通信协议——TCP/IP（传输控制协议/互联网协议），并创建了第一个连接不同机构的实验性网络，互联网由此诞生。

很快，TCP/IP协议就将几千台（后来又增至几百万台）计算机连接起来，并相互交换文件和文档。网络浏览器被开发出来后，即使是非专业用户也可以便捷地访问万维网。网络浏览器是一种显示文档的程序，易于跟踪超文本链接。第一个得到广泛使用的网络浏览器Mosaic（马赛克），是伊利诺伊大学厄巴纳－香槟分校的一个由马克·安德森带领的团队于1993年发布的，它代表了万维网的一个转折点。互联网的成长和万维网的普及让全世界大吃一惊，可以说，没有任何一种现象像互联网这样迅速且彻底地改变了世界各地的文化和日常生活。

图3－8展示了自互联网诞生以来网络用户数量的变化情况。在20年里，网络用户由1993年的少数人增加为全世界的大部分人口。其他任何技术（除了一些基于万维网开发的技术）可能都不曾像万维网这样迅速地改变了世界。

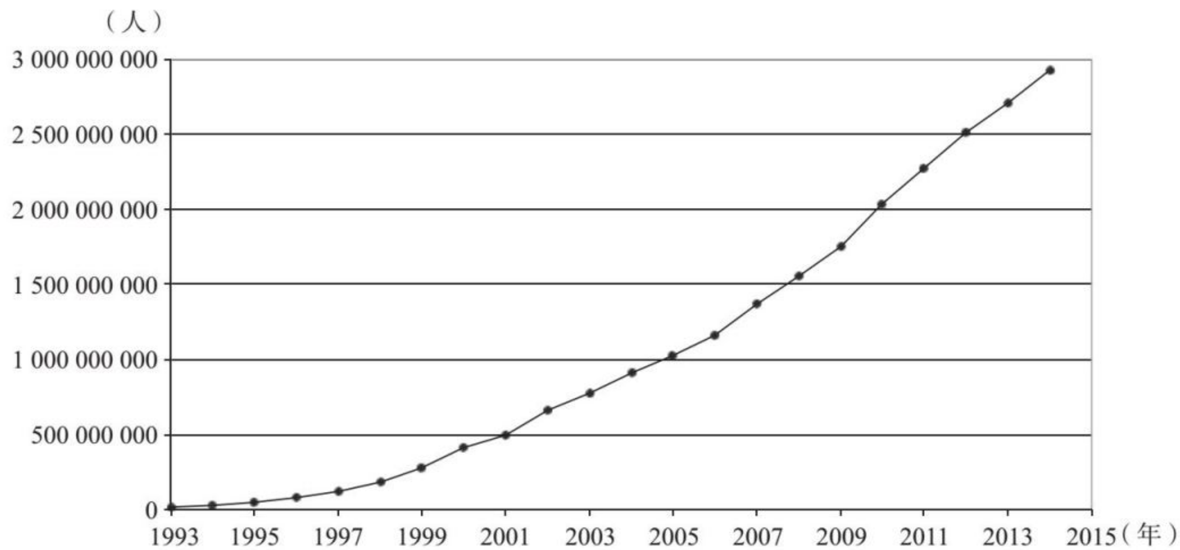


图3－8 互联网用户数量的变化情况

最初，万维网允许用户访问存储在服务器（由专业人员维护的大型计算机，可同时为许多用户服务）中的内容，它们主要是由专业人员或高级用户开发的。但是，随着更多的用户友好应用和界面被开发出来，几乎任何用户都可以创建内容，包括文本、博客、图片和电

影。这催生了Web 2.0（万维网2.0版），并推动了网络效用的二次增长。例如，如果用户数量增加一倍，并且他们都对丰富万维网的内容做出了贡献，那么所有用户可以使用的知识量将变为原来的4倍。网络效用的二次增长又进一步推动了新应用程序的开发和万维网的使用，其中有许多都是几年前预料不到的。

仅凭万维网上可以便捷获取的海量信息，就足以改变世界。我们中的许多人应该还记得，曾几何时我们需要付出多大的努力，才能查询到标准百科全书中没有收录的专业问题。今天，互联网上的一次简单搜索，即使是最晦涩的话题，也能得到成百上千页的反馈。随着Web 2.0的出现，存储在互联网上的有条理的信息数量呈爆炸性增长。在我撰写本书期间，在线百科全书——维基百科的英文版包括460多万篇文章和10多亿个单词，单词数量是有史以来出版的最大的英语百科全书——《不列颠百科全书》的30多倍。维基百科文章数量的增长情况（图3－9）表现为一条加速曲线，尽管随着维基百科逐渐覆盖了与大量人群相关的大多数知识，它的增长呈现出减速的趋势。

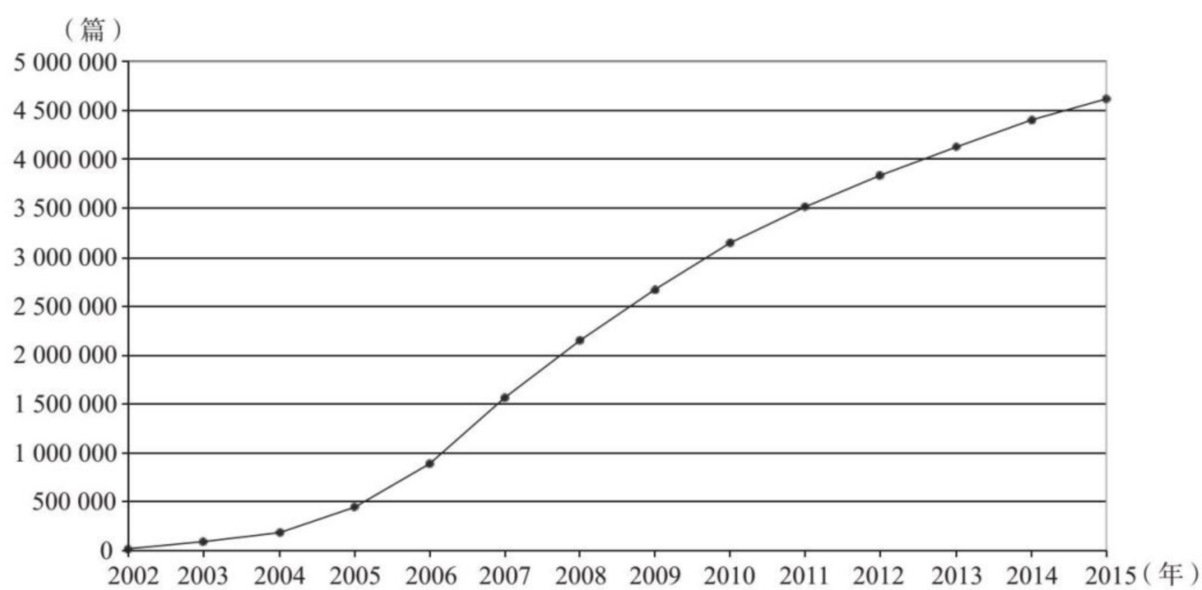


图3－9 维基百科文章数量的变化情况

大量用户使不断增长的社群进一步增值，带来了网络效用的二次增长，维基百科只是众多这类案例中的一个。其他著名的例子还包括

视频网站YouTube（用户可以上传视频）、图片分享网站Flickr和Instagram（用户可以上传照片），以及一系列社交网站，其中最流行的是脸书。从领英之类的职业网站到推特之类的允许用户使用计算机或手机随时随地发布碎片化信息的网站，各种各样的专业网站几乎可以迎合所有人的品位或信念。[最新电子书免费分享社群，群主V信1107308023 添加备注电子书]

电子商务——利用万维网做营销、销售和运输实体商品——正在以10年前人们预测不到的方式改变着世界经济。现在，在线订购书籍、音乐、食物和许多其他商品并送货上门，已经变得司空见惯。2015年，亚马逊网站成为世界上最有价值的零售商，超过了最大的实体店。尽管如此，亚马逊仍然保留了存储和运输所售商品的实体设施。优步和爱彼迎的变化更加彻底，它们开始利用完全虚拟化的基础设施提供服务（分别是交通和住宿方面的服务），并对那些以实体设施的形式（出租车和酒店）提供这些服务的公司构成了严重威胁。

网络游戏是计算机和互联网对人们的日常生活方式产生深远影响的又一个例子。一些大型多人在线角色扮演游戏（MMORPG）已经积累了大量注册用户，他们在虚拟游戏世界中进行交互。这些通过互联网相互联系的玩家在玩游戏的同时，按照虚拟世界的规则建立起长期的关系并进行互动。在我撰写本书期间，最受欢迎的游戏拥有几百万注册用户，相当于一个中等规模国家的人口数量。仅存在于网络游戏虚拟世界中的商品，成为现实世界中的常见交易商品，有些甚至是高价商品。

另一种形式的交互可能是未来事物的一个预兆，即在虚拟世界中，只要技术允许，人们就可以尽可能接近现实世界的方式生活、工作、买卖房产和从事其他活动。2003年推出的网络游戏《第二人生》，可能是这类虚拟世界模拟程序中最著名的一个。《第二人生》拥有与现实经济平行的经济体，它使用的虚拟货币可以像传统货币一样兑换。这个虚拟世界中的公民可以开展许多与现实世界平行的活

动。服务条款保证用户拥有他们创建内容的版权，系统还提供一些简单设施，便于用户管理数字版权。目前，用户界面在真实感方面还存在一定的局限性，因为在与虚拟世界交互时，用户只能使用基于键盘的界面和较低分辨率的计算机生成图像。尽管《第二人生》的发展速度比较缓慢，但它现在拥有约100万个固定用户。

以上是关于互联网技术对日常生活影响的非常概要的介绍，肯定极其不全面。关于这方面的更多信息你们可以在万维网上找到。不过，即使是像这样粗略的描述也足以清楚地表明，几百万用户使用的在线服务在几年前根本不存在。

当今社会的一个明确特征是，它极度依赖信息和通信技术。大约60年前，IBM推出了它的第一台电子计算机——IBM701。当时，只有一小部分经济依赖数字技术；电话网络对经济而言非常重要，但它们建立在模拟技术的基础之上，依赖数字计算机的经济产出少到几乎可以忽略不计。

如今，数字技术是经济的不可或缺的组成部分，以至于我们很难甚至不可能计算出它们对经济产出的贡献率。虽然计算机设备制造商、软件开发商和数字产品制造商创造的总价值，的确有可能被计算出来（其中，计算出数字产品制造商创造的价值的可能性要小一些），但数字技术与大部分人口的每一项活动息息相关，以至于我们不可能计算出这些技术对整体经济的间接贡献。若干研究已经在解决这个问题，但仍未给出数字技术对经济产出所做贡献的具体价值。

不过，显而易见的是，数字技术在经济中占据的比例不断增大，这一比例从60年前的零稳步上升至今天的显著份额。数字技术对美国GDP（国内生产总值）的直接贡献超过1万亿美元，占比超过7%，而且20年来，这一占比的增速达到4%，在历史上任何其他行业的增速都无法与之相提并论。然而，这并未考虑数字技术对日常生活的所有影响，如果将其全部计算在内，数字技术的GDP占比将会高得多。

我们没有理由认为数字技术在经济领域的重要性将会减弱，也没有理由认为它的增长速度将会降至一个更合理的数值。相反，有充分的证据表明，在未来几十年里，这些技术将在经济产出中占据更大的比例。数字技术的GDP占比看似会在未来某个时刻停止增长，毕竟一些重要的需求（例如食品、住房、交通、服装和能源）是无法通过数字技术满足的，而且这些需求在整体经济活动中的占比肯定有一个最小值。例如，我们可以保守地认为，整体经济产出中的某个固定比例（比如，50%）必定要用于满足实际的物质需求，因为计算机、手机和其他数字设备是无法满足这类需求的。但是，这是基于整体经济产出将在某一时刻停滞不前的想法产生的一种错觉——这种情况从未发生，近期也不太可能发生。尽管基本需求必须予以满足（至少在相当长的一段时间内如此），但基于数字技术的新服务和新产品的前景从根本上说是无限的。由于数字技术对经济增长的贡献大于其他技术和产品，因此我们可以预期，在未来的某个时刻，纯粹的数字商品将在经济产出中占据更大的比例。事实上，数字经济的整体贡献没有明显的上限。与实物商品不同，数字商品不受自然资源（比如原材料、土地或水）可用性的限制。计算机技术的迅速发展，使我们有可能在不需要耗费额外资源的前提下部署新产品和新服务，而不仅限于现有的计算平台。随着计算机变得越来越节能，额外的能源需求可能微乎其微，甚至为零。

这在历史上并不是什么新鲜事。直到几百年前，几乎所有的家庭收入都被用来满足食物、住房等基本需求。随着技术革命的推进，一小部分家庭收入流向了不太基本但仍然非常重要的东西，比如交通和服装。家庭收入不断流向在我们看来不那么重要的商品和服务，这个持续性变化与早在农业出现之时就已经开始的趋势一脉相承。

有人可能认为，在某个时刻，流向数字商品和服务的收入比例将停止增长，原因很简单，即人们没有更多的时间或资源来使用这些技术。一个人一天能花多少个小时看数字电视、浏览网页或发短信呢？肯定不超过24个小时，而且在大多数情况下远少于24个小时。毕竟，

吃饭和睡觉肯定要占用一部分时间。但是，这种想法忽略了一个事实：数字经济可以在没有人类直接干预的情况下创造价值。正如我们将在第11章中看到的那样，数字智能体可以代表人类或者企业，创造数字商品和服务，供其他人（包括其他数字实体）消费。终有一天，人类直接活动在整体经济产出中的占比将显著低于100%。在很大程度上，今天的情况已然如此。事实上，作为我们的重要经济支柱的数字服务，是由计算机完成的，人类在其中并没有发挥不可或缺的辅助作用。不过，现在这些服务是代表某些公司提供的，归根结底它们是由人类所有者或股东控制的。最终，计算经济贡献的标准计算会将公司贡献的价值归功于公司的所有者。

从这个意义上说，今天的所有经济产出都是人类活动的结果。的确，在许多情况下，所有权很难追踪，因为有的公司归属于其他公司。不过，一家公司的最终所有者肯定是某个人或某群人。因此，尽管有些经济产出实际上是由非常自主的，有时甚至是非常智能的系统创造的，但这些经济产出仍然会被归功于人类。除非有一天，某些计算主体被赋予与人类或公司相当的人格权，并可被视为商品或服务的最终生产者，这种情况才有可能改观。只有到那时我们才会改变看法，不再把世界经济视为人类活动的产物。

但在这一天到来之前，我们必须更深入地理解为什么计算机拥有以往的任何技术都不具备的颠覆性潜力。在历史上，计算机的出现早于晶体管，与电的发现差不多在同一时期。尽管在电子设备问世之后，人们才建造出真正可用的计算机，但计算理论本身有一个相似却又独立的发展史。

---

1. 1平方英尺≈ 0.09平方米。——译者注
2. 1加仑≈ 3.8升。——译者注

第4章

通用计算机

现在，计算机无处不在，所以我们很难想象没有计算机的时代是什么样子。“computer”这个单词的最早使用时间可能是17世纪，指进行计算或运算的人。直到20世纪中叶，它的意思才发生变化，指执行一系列简单计算（可编程）的人或物。16~17世纪，少量机械计算设备被建造出来，但没有一个是可编程的。查尔斯·巴贝奇率先设计了一台完全可编程的机械计算机，并完成了部分建造工作，他称之为分析机。

分析机

分析机的前身是差分机，后者是巴贝奇设计并建造的一种计算器，可以自动完成天文和数学表格的计算工作。它的运行建立在数字按列排布的表格法的基础之上，这台机器在每一列存储一个十进制数字，将第n+1列中的一个单元格的值与第n列中的一个单元格的值相加，它们的和就是第n列中的下一行的值。利用这类重复计算，可以算出多项式的值。

为了理解差分机的工作原理，我们需要看一个具体的实例。例如，考虑由多项式$P(x)=x^2+2x+2$定义的函数。列出x的前几个值和连续差分，就会得到如表4－1所示的表格。从表中我们可以清楚地看到，这个多项式的二阶差分是常数，这是二次多项式的一般性质。依据这个性质，我们不需要进行任何乘法运算，就可以计算出多项式在任意点的值。例如，当x=7时，多项式的值为65，即2（D2列中的常数）与D1列中的13相加，再加上P(x)列中的50。

表4–1 差分机使用的差分表

x	P(x)	D1	D2
1	5		
2	10	5	
3	17	7	2
4	26	9	2
5	37	11	2
6	50	13	2

在为差分机编程时，需要设定各列的初始值。一开始，先确定P(x)列的前几行（行数等于多项式的次数加1）的值，也就是多项式的值。D1列中的值是通过计算多项式的连续点之间的差分得到的，以人工计算的方式减去前一列中连续各行的值，就可以得出下一列的初始值。然后，差分机从最后一列开始，将每列最后一个单元格的值与前一列最后一个单元格的值相加，一直加到P(x)列，即可计算出x取下一个值时多项式的值。这样一来，差分机就可以完成任意行数的计算工作。

差分机可以进行重复计算，但它不能通过编程的方式进行任意序列的重复计算，因此它不是可编程计算机。巴贝奇的第二个设计，也就是目标更加远大的分析机，通常被视为第一台真正的可编程计算机。

在建造这台分析机时，巴贝奇构思了详细的计划，并完成了部分建造工作，如图4－1所示。程序和数据以穿孔卡片的方式输入这台机器。分析机利用打印机输出结果，但也可以使用其他输出设备，其中包括卡片穿孔机，它还可被用于生成接下来的输入。

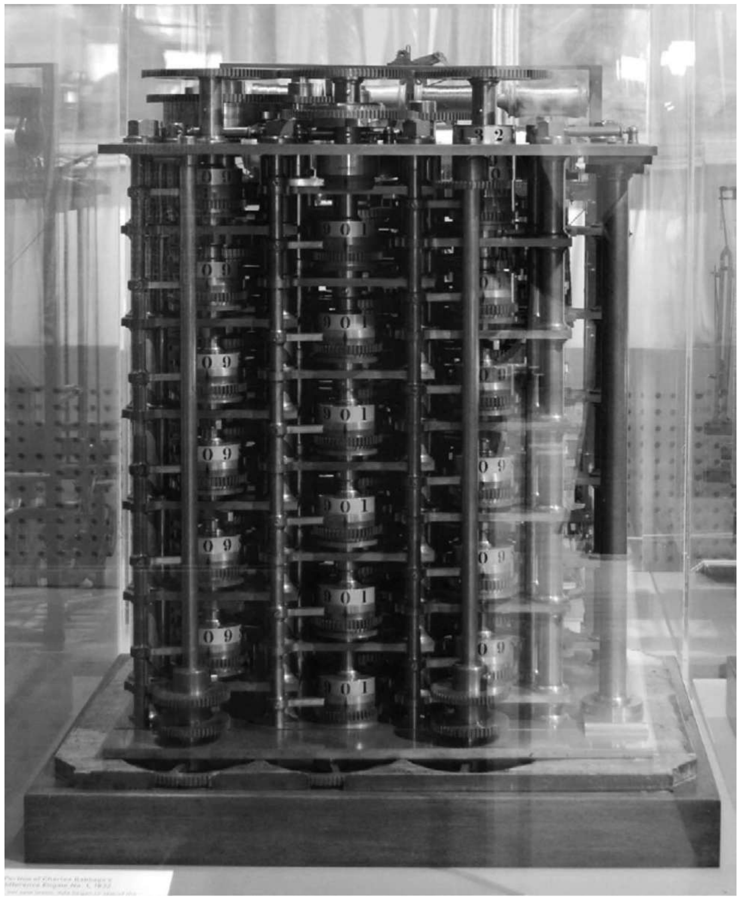


图4－1 现陈列于伦敦科学博物馆的分析机的一部分

这台机器可以存储1 000个50位的十进制数，其中央处理器可进行4种算术运算和比较。分析机的基本运算是加法，这种运算和其他基本运算——减法、乘法和除法——都要在“磨坊”（mill）里进行。那里有一个旋转凸轮系统，在钟形曲柄的作用或驱动下，完成这些运算。

分析机用户使用的编程语言与现在的中央处理器使用的汇编语言同样抽象。汇编语言的作用通常是，帮助程序员指定不同存储位置之间的基本算术和逻辑运算序列，以及这些运算的完成顺序。分析机的编程语言包括控制指令，比如循环（指示计算机重复特定操作）和条件分支（指示计算机根据某个已经算出的值去执行下一条指令）。

如果分析机被成功建造出来，它就应该是一台真正的通用计算机。1842年，路易吉·梅纳布雷亚写了一本关于分析机的法语版书

籍，第二年，这本书被翻译成英文，阿达·洛芙莱斯（拜伦勋爵的女儿）还对它进行了注解。洛芙莱斯早在10年前就对分析机产生了兴趣，她在注释（篇幅比梅纳布雷亚对分析机的描述长好几倍）中指出，这台机器不仅可以处理数字，还可以处理其他符号：

> 如果对象之间的基本相互关系可以用抽象运算科学之间的关系来表示，并且这些对象易于改变以适应分析机的运算符号和机制，那么，除了数字以外，它还可以作用于其他事物。例如，假设在和声学与音乐作曲学中，音调之间的基本关系可以采用这种表达方式，那么分析机或许可以创作出任意复杂程度、任意音域的科学而精妙的音乐作品。

她也提出了一种利用分析机计算伯努利数的方法，被视为对计算机程序的最早精确描述。由于这些贡献，阿达·洛芙莱斯被公认为第一个计算机程序员，也是第一个认识到计算机不仅能够处理数字，还能够处理通用符号序列的人。编程语言Ada（阿达）就是以她的名字命名的。

技术和实际困难致使查尔斯·巴贝奇未能制造出一台可以工作的分析机，尽管在1871年去世之前他努力完成了一小部分的组装工作。1992年，工作版本的巴贝奇分析机终于建成，目前陈列在伦敦科学博物馆中。

图灵机和计算机

在巴贝奇之后，计算机的下一次根本性进步要归功于科学家和数学家阿兰·图灵，他提出了符号处理装置的一个抽象模型，用于研究算法、程序和计算的性质。

图灵机的设计初衷并不是建造实用的计算机器，而是计算设备的抽象模型。确定型图灵机（本章简称为图灵机）包含以下几个部分：

- 一条无限长的纸带，被分成一个个单元格。每个单元格中有一个符号，这些符号来自某个有限的字母表。该字母表包含一个特殊的空格符号，以及一个或多个其他符号。尚未写入的单元格中填充着空格符号。
- 一个带头，可以读写纸带上的符号，每完成一次读写，就将纸带向左或向右移动一个单元格。
- 一个状态寄存器，用于存储图灵机的当前状态。图灵机有两种特殊状态，即接受状态和拒绝状态。当进入其中一种状态时，图灵机会停止运行，从而确定输入被接受还是被拒绝。
- 一幅状态转换图，用于指定转换函数。状态转换图可以根据图灵机的每种配置，描述带头应该完成的动作（是擦除一个符号还是写入一个新符号，并向左或向右移动），以及图灵机接下来的状态。

图4－2是图灵机示意图，其中状态转换图描述了有限状态控制器的工作原理。状态转换图有4种状态（用圆圈表示），各种状态之间有若干种转换（用连接圆圈的箭头表示）。控制器根据指定的转换函数

从一种状态转变为另一种状态，每一种转换都是由一对输入/输出规定的，输入和输出之间用“斜杠”（/）分隔。但在图4－2所示的简单情况中，图灵机没有输入，因此图中只显示了输出。图4－4展示的是一种略微复杂的情况，每次转换都标记了一对输入/输出。

图4－2依据的是图灵在他1937年发表的论文中提出的第一个例子。这台机器有4种状态：初始状态b（圆圈加粗以示强调）和其他三种状态c、e和f。因为这台机器没有接受状态和拒绝状态，因此它只是不停地运行，并在纸带上写出无数的“0”，“ ”，“1”，“0”，“ ”，“1”…。它的行为很容易理解。图灵机从状态b启动，此时纸带上填满了空格符号。状态转换图告诉这台机器写入一个0（P0），向右移动（R），并转换为状态c。在状态c时，这台机器被告知向右移动（R），不写入任何符号，在纸带上留下一个空格符号。接下来，图灵机以相同的方式继续运行下去。在图4－2中，控制器处于状态f，机器运行并进入状态b，带头向右移动。虽然这个例子非常简单，图灵机执行的也是一项非常简单的任务，但它也可以执行更复杂的任务。例如，图灵机可以通过编程算出写在纸带上的任意两个二进制数的和，并将结果写回到纸带上。事实上，图灵机可以通过编程计算任何数字或执行任何符号处理任务，前提是这些任务可以描述为一系列步骤（即算法）。它可以计算多项式在某一点上的值，判断一个方程是否有整数解，以及确定一个图像（纸带上描述的图像）是否包含另一个图像（纸带上的另一个位置描述的图像）。

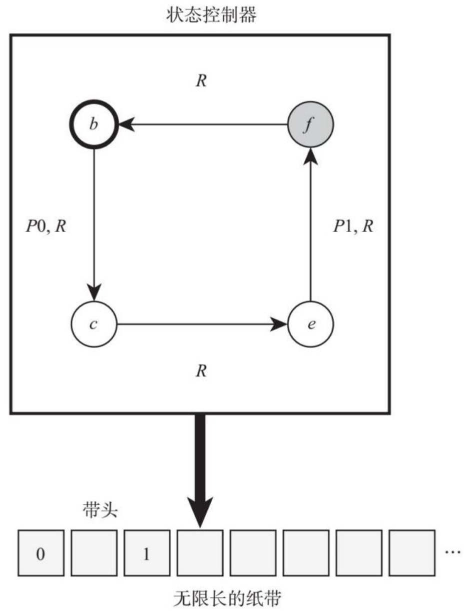


图4-2 一台会在纸带上写入无数个0和1（中间有空格）的图灵机

特别是，一台足够强大的图灵机可以模拟另一台图灵机。这种特殊形式的模拟被称作仿真，如果一个系统模拟另一个完整系统的行为，并且输出的行为达到无法区分的程度，我们就称之为仿真。

如果一台图灵机可被另一台图灵机仿真，我们就有必要明确说明“足够强大”的含义。显而易见，有些图灵机的功能还不足以执行复杂任务，比如模拟另一台机器。例如，上述例子中的图灵机除了写入0和1之外，其他什么都做不了。但是，有些图灵机非常复杂，足以模拟其他任何一台图灵机。如果图灵机A把对图灵机B进行的恰当描述（表示为<B>），以及图灵机B的输入一起写在纸带上，我们就说图灵机A可以模拟图灵机B，它是一台通用图灵机。如果图灵机A想要模拟图灵机B，就要将描述图灵机B的字符串写到图灵机A的纸带上，符号<B>的作用是标示这个字符串。

通用图灵机的概念为我们思考计算能力提供了一个新思路。例如，如果我们想弄清楚图灵机可以计算什么，我们不需要想象所有可能的图灵机。我们只需要想象一台通用图灵机，然后把关于图灵机的所有可能的描述（可能还有它们的输入）都输入到这台通用图灵机

中。之后，当我们为了评估它们的计算能力而计数和枚举图灵机时，这种推理过程将会助我们一臂之力。

图灵机是一种关于计算机的抽象数学模型，在理论计算科学中的主要作用是阐述并解决问题。图灵机的设计初衷并不是建造出一台真正可用的实物计算机。随着真正的电子计算机的出现，更切实可行的计算机模型已经变得必不可少。最基本的想法——在很大程度上源于图灵的研究成果——是存储程序计算机，这种计算机利用存储器而不是纸带存储代表特定程序的指令序列。1946年，图灵在向英国国家物理实验室执行委员会呈交的一篇论文中，提出了第一个关于存储程序计算机的合乎逻辑的完整设计方案，他把这种设备称为自动计算机。但是，熟知图灵理论研究的约翰·冯·诺依曼设计的EDVAC（离散变量自动电子计算机），被公认为存储程序计算机的第一个实际可行的方案。

约翰·冯·诺依曼出生于匈牙利，6岁时就能心算8位数的除法，是一个天才儿童。1930年，他受邀成为普林斯顿大学的客座教授，其间在数学、物理学、经济学和计算等领域都取得了辉煌的成就。他与理论物理学家爱德华·泰勒、斯塔尼斯拉夫·乌拉姆共同提出的一些概念，最终促成了曼哈顿计划和原子弹的研制。

冯·诺依曼做出的诸多科学贡献之一，就是所谓的冯·诺依曼体系结构。在《EDVAC报告书的第一份草案》（*First Draft of a Report on the EDVAC*）中，他提出了一种数字计算机的体系结构，包括中央处理器、控制器、中央存储器、外部海量存储器和输入/输出设备。中央处理器包含控制器和算术逻辑部件。算术逻辑部件对存储在中央处理器的寄存器中的值进行算术逻辑运算。控制器中有一个程序计数器，它指向存储器中包含当前指令的特定点，并将执行的操作代码加载到指令寄存器（中央处理器中的一个寄存器）中。用来存储数据和指令的存储器，可以通过中央处理器中的寄存器进行访问。外部海量存储器的作用是永久地存储指令和数据，以备将来之用，而输入

和输出设备的作用是让计算机能与外部世界通信。存储器和外部海量存储器取代了图灵机的纸带，中央处理器的内部结构则使存储器和固定存储器中的信息处理效率更高。

在所谓的冯·诺依曼体系结构中，加载程序指令以及在中央处理器和存储器之间移动工作数据，使用的是同一条路径，这一限制可能会减慢计算机的运行速度。所谓的哈佛结构（得名于采用全继电器的哈佛马克I号计算机），在加载程序指令以及在中央处理器和存储器之间移动工作数据时，使用的是不同的路径（或总线）。有些现代计算机（包括专门用于处理数字信号的计算机）使用的是哈佛结构，还有一些计算机使用的是哈佛结构和冯·诺依曼体系结构的组合。

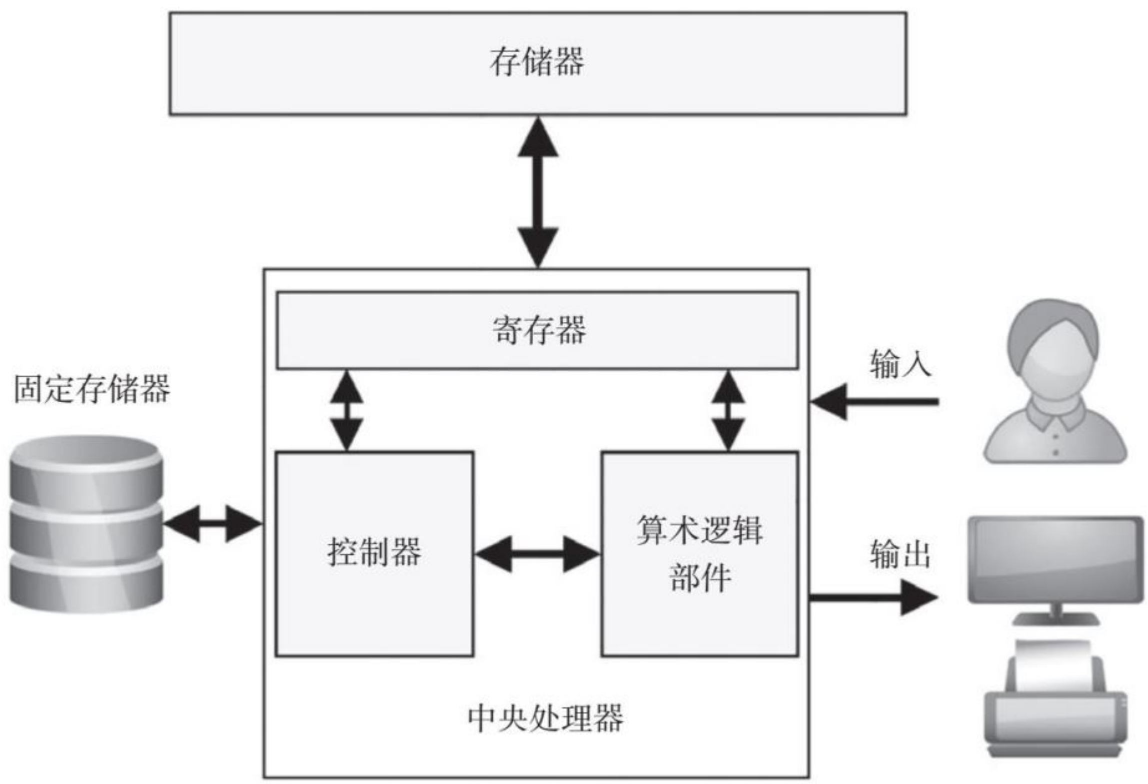


图4－3 存储程序计算机的冯·诺依曼体系结构

现代计算机有多条总线和多个中央处理器，并运用复杂的机制来确保不同的中央处理器都可以访问可用存储器（可用存储器有许多种不同的排布方式）。但是，所有现代计算机的工作方式在本质上是相

同的，都是根据阿兰・图灵和约翰・冯・诺依曼的存储程序计算机设计方案建造而成的高速计算机。

可计算性和无穷悖论

与存储程序计算机等更加复杂的计算机模型相比，图灵机的优点是，关于它的性质的数学分析更加简单直观。虽然图灵机看起来非常抽象和不实用，但图灵证明在某些假设条件下，这样的机器能够执行任何可能的计算。这意味着图灵机和其他计算设备一样强大，换句话说，图灵机是一台通用计算机，它可以计算任何可计算的东西。

我们现在知道，几乎所有已经存在的计算模型和语言都具有图灵意义上的等价性，即可以用其中一种编程的东西，换成另一种也可以做到。因此，什么可以计算，什么不可以计算，就有了一个非常清晰的定义。这样的结果看起来令人惊讶，所以需要做更深入的解释。如此简单的一台机器，只使用一条纸带作为存储器，怎么可能跟其他计算机一样强大呢？“任何可计算的东西”到底是什么意思呢？有不能计算的东西吗？为了进一步讨论什么可以计算而什么不能计算，我们必须明确无误地回答这些问题。为此，我们需要使用语言的概念。

语言是指源自给定字母表的一组符号序列。例如，根据这个定义，英语是能够满足特定句法和语义规则的所有可能的字母符号序列。再例如，考虑语言L^1 ={0，00，000，0000，00000，…}，它由所有只包含0的字符串组成。当然，这样的字符串有无穷多个，因此，这种语言有无穷多个字符串。

假设存在一台图灵机，当我们在纸带上写入一个字符串时，如果该字符串属于语言L，那么机器进入接受状态并停机，如果该字符串不属于语言L，则机器进入拒绝状态并停机，据此我们说语言L是可判定的。由于这样的图灵机确实存在（图4－4中的图灵机可进行两种不同

的计算），因此语言L_1是可判定的。图4－4的两个部分展示了这台图灵机进行的两种计算，以及它在输入0001［图4－4（a）］和000［图4－4（b）］后的状态。在图4－4（a）中，图灵机拒绝了字符串，控制器在看到纸带的第4个单元格中的1后进入拒绝状态。在图4－4（b）中，图灵机在看到纸带上0的字符串后面的一个空格后，接受了这个字符串。

如果存在一台图灵机，当处理属于语言L的字符串时进入接受状态并停机，而当字符串不属于语言L时无法进入拒绝状态并停机，我们就说语言L是可识别的。因此，可判定语言比可识别语言“简单”，因为对于可判定语言，图灵机可以肯定地回答“这个字符串是否属于L？”的问题，但如果L是可识别语言，那么只在答案是肯定的情况下，图灵机才能回答这个问题。

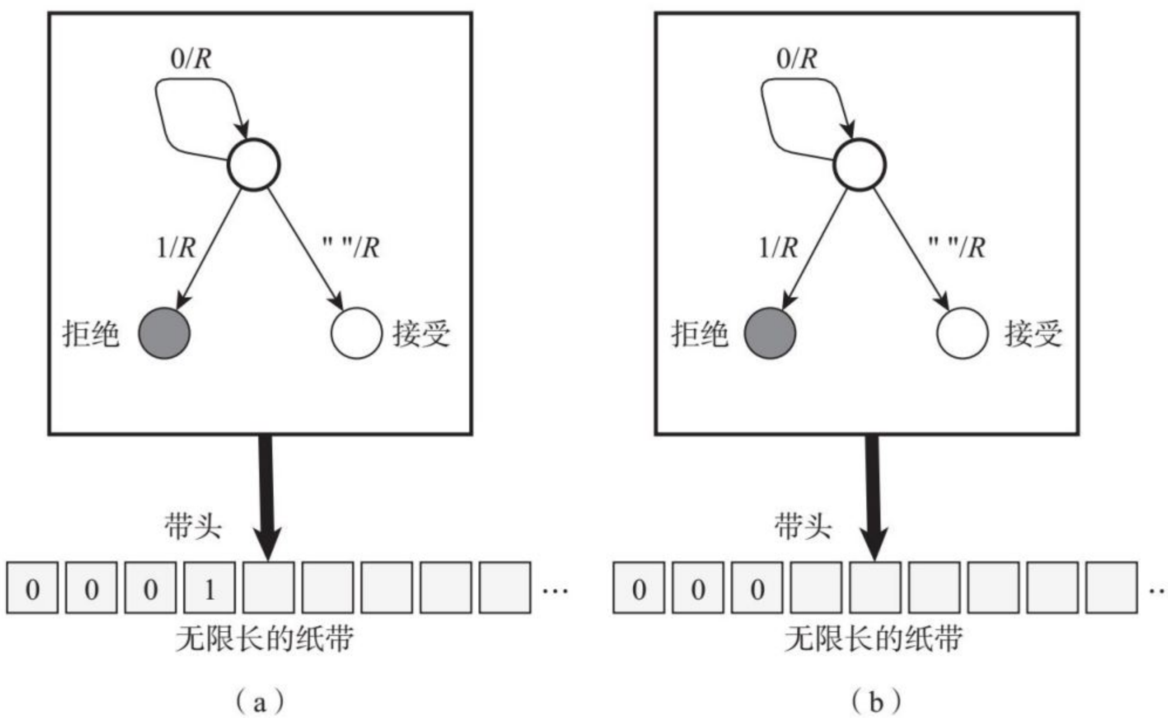


图4－4 接受语言L_1的图灵机

语言和问题之间有什么关系呢？我们在这里使用的“问题”一词，指的是有特定答案的问题。如果我们将分析限定在答案只能为

“是”或“否”的问题上，就有可能确定问题与语言之间的对应关系。每个问题实例都会被编码成一个字符串，只有与答案为“是”的问题实例相对应的字符串，才能定义与这个问题相对应的语言。而其他字符串要么与问题实例不对应，要么与答案为“否”的问题实例相对应。与可判定语言相对应的问题被称为可判定问题，否则，我们就说这个问题是不可判定或者不可计算的。

更务实的说法是，不可判定问题指所有算法都无法保证找到答案的问题。这意味着，计算机在有限的时间内执行任何固定指令，都不能保证找到答案。发现一些问题不可判定的事实，是阿兰·图灵对计算理论做出的主要贡献之一。阿隆佐·丘奇运用不同的形式主义，得出了不一样的可计算性定义（现在，人们发现丘奇的定义与图灵的定义其实是一样的）。库尔特·哥德尔的不完全性定理认为，数学真理的存在不能通过固定的公理来证明或推导，归根结底这等同于图灵得出的存在不可判定或不可计算问题的结论。

对我们而言，理解为什么不可判定的问题必然存在，是至关重要的。因此，我们必须更深入地研究可计算性和无穷的概念。为了证明不可判定的问题必然存在，我们需要进一步熟悉无穷的概念，并且知道并非所有无穷集都拥有相同数量的元素。考虑自然数集{1,2,3,…}和偶数集{2,4,6,…}，它们都是无穷基数。它们的基数相等吗？或者说，它们的元素个数相同吗？有人可能会说自然数比偶数多，因为所有的偶数都是自然数，而自然数还包括所有奇数。但是，如表4－2所示，我们可以在自然数集和偶数集之间建立一一对应关系。

我们现在有充分的理由认为，如果两个集合的元素之间可以建立一一对应关系，这两个集合就拥有相同数量的元素，即它们的基数相等。这正是数学家格奥尔格·康托尔在1874年给基数下的数学定义，当集合中的元素个数为无穷时，它是我们可以合理使用的唯一定义。

表4–2 自然数和偶数的一一对应关系列表

自然数	偶数
1	2
2	4
3	6
4	8
5	10
…	…

因此，偶数和自然数的个数一样多。现在，考虑所有正有理数的集合，这些数可以写成两个正整数之比的形式，包括0.5(1/2)，0.25(1/4)，0.3333…(1/3)和1.428 571 428 571…(10/7)等无穷多个数字。正有理数肯定比自然数多，因为直觉告诉我们，任意两个整数之间存在无穷多个有理数。事实上，情况甚至更加极端，因为任选两个有理数，它们之间也存在无穷多个有理数。

遗憾的是，直觉再一次让我们失望了。有理数集的基数并不比整数集的基数大。使用一种更聪明的结构，就有可能列举出所有有理数。如果我们可以建立包含所有有理数的有序表，它们就可以与自然数形成一一对应关系，从而证明有理数集与自然数集的基数相等。

为了列举出所有正有理数，我们先来考虑分子与分母之和为2的有理数，然后是分子与分母之和为3的有理数，之后是分子与分母之和为4的有理数，以此类推。需要注意的是，相同的数字不能列举两次。例如，如果列表中已经包含了1/2，就不应该将2/4列入其中。通过这种方法列出所有正有理数，最后会得到表4－3。我们应该清楚的是，每个正有理数都会出现在这个列表的某个地方，而且有理数和自然数之间存在一一对应关系。这证明它们具有相同的基数。

表4–3 自然数与有理数的一一对应关系列表

自然数	有理数
1	1/1
2	1/2
3	2/1
4	1/3
5	3/1
6	1/4
7	2/3
…	…

与偶数列表一样，正有理数集也是可数集，因为它可以与自然数形成一一对应关系。利用类似的技巧，我们似乎可以为所有无穷集创建列表。只要集合的所有元素能以某种特定的顺序列出，这个集合就不会比自然数集大。这意味着所有无穷集都具有相同的基数，即自然数集的基数$\aleph_0$（阿列夫零）。遗憾的是，事实并非如此，因为正如康托尔证明的那样，有些集合的元素无法建立有序表。我们可以通过一个简单的方法，证明这些集合的基数比自然数集大。

假设我们可以建立一个0到1之间的实数列表。它类似于表4－4，其中α_i、β_i、γ_i等分别表示一个特定的数字。现在，考虑数字0.αβγδε…，其中$\alpha \neq \alpha_1$，$\beta \neq \beta_2$，$\gamma \neq \gamma_3$，以此类推。例如，如果$\alpha_1=5$，我们就令这个数字中的α等于6（或者不等于5的任何数），否则，我们令α等于5。同样，如果$\beta_2=5$，就令β=6，否则令β=5。以此类推。数字0.αβγδε…必然存在，但它很显然不在列表中。因此，并非所有0到1之间的实数都在这个列表中。从这个推理过程中只能得出一个结论：我们无法建立一个完整的实数列表，哪怕

只是0和1之间的实数。由此可见，这些数字集的基数大于自然数集的基数。

表4–4 自然数和实数之间的一一对应关系的假设列表

自然数	实数
1	$0.\alpha_1\beta_1\gamma_1\delta_1\varepsilon_1\cdots$
2	$0.\alpha_2\beta_2\gamma_2\delta_2\varepsilon_2\cdots$
3	$0.\alpha_3\beta_3\gamma_3\delta_3\varepsilon_3\cdots$
4	$0.\alpha_4\beta_4\gamma_4\delta_4\varepsilon_4\cdots$
…	…

现在我们来考虑某个固定字母表Σ，以及其中的符号可以构成的所有可能的有限字符串的集合。我们称这种语言为Σ*。这个集合是可数集，因为我们可以用列举所有有理数的方法列举出所有字符串。例如，我们可以从长度为0的字符串（只有一个，即空串，通常用ε表示）开始，然后列出长度为1，2，…的字符串，按照这种顺序，我们可以得到类似于(s_1，s_2，s_3，s_4,…)的列表。在字母表Σ的基础上定义的所有语言L都是Σ*的子集，并且可以通过由0和1构成的（无限）字符串，识别出列表中的哪些字符串属于L。例如，如果Σ={0,1}，Σ*中的字符串列表就会类似于(ε，0，1，00，01，10，11，000，001，…)。举个例子，考虑所有字符串的开头均为0的语言L_2，以及所有字符串均有偶数个0的语言L_3。这两种语言（以及在Σ*基础上定义的所有语言）都可以用一个由0和1构成的唯一（无限）字符串来定义，如果Σ*的字符串列表中的一个字符串属于该语言，则无限字符串中与之对应的字符为1，否则为0。图4–5展示了这种对应关系。

在图4–5的上方，我们按照某种特定顺序列举了可以用字母表Σ写出的所有字符串。顺序本身并不重要，只要它保持不变即可。在L_2

方框的底部，我们列出了属于L_2的所有字符串。在本例中，语言L_2包含所有开头为0的字符串。在L_2方框的顶部是一个能定义L_2的无限字符串，由0和1组成。L_3方框的格式亦如此。记住，所有包含偶数（包括0）个0的字符串都属于语言L_3。它们构成的集合位于方框底部，定义L_3的由0和1构成的无限字符串则位于方框顶部。显然，在字母表Σ基础上定义的语言与由0和1构成的无限字符串之间存在一一对应的关系。现在，事实证明，由0和1构成的所有无限字符串的集合是不可数集。我们无法构建这个集合的元素列表，出于同样的原因，我们也无法构建实数集的元素列表。运用我们在证明实数集不可数时使用的对角线论证法，可以得出同样的结论。这意味着在字母表Σ基础上定义的所有语言构成的集合是不可数的。

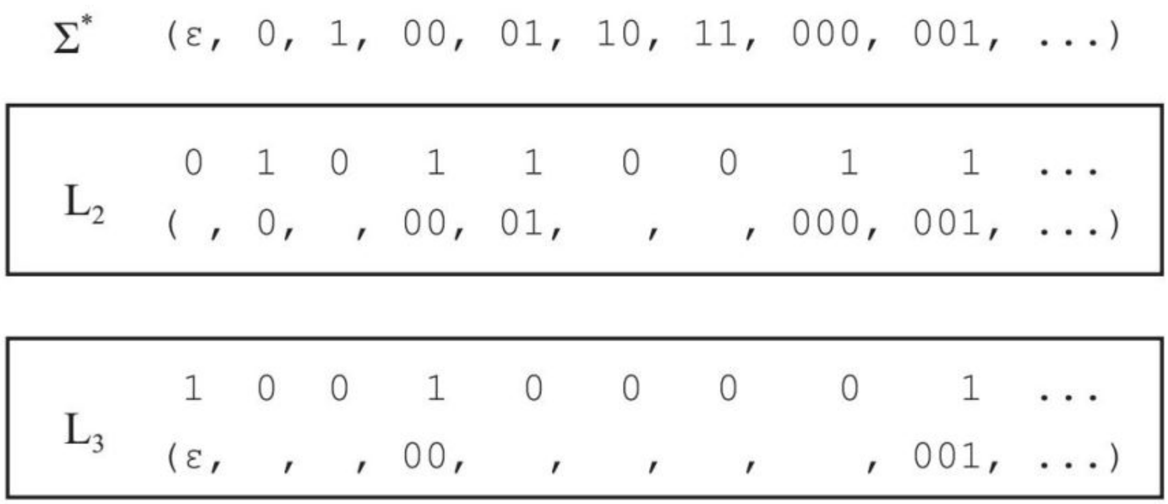


图4－5 由0和1构成的无限字符串与在字母表基础上定义的语言之间的等价关系示意图

研究表明，在字母表Σ基础上运行的所有图灵机构成的集合是可数的。我们可以构建包含所有图灵机的（无限）列表，先列出状态值为0的图灵机，再列出状态值为1的图灵机，然后列出状态值为2的图灵机，依此类推。或者，我们可以想象有一台通用图灵机，只要输入相关描述，它就可以仿真任何图灵机。关于这些图灵机的描述是一个固定的符号集，它们可以枚举，因此我们能构建包含所有图灵机的列表，它类似于{TM_1， TM_2， TM_3， TM_4,…}。

现在，我们知道建立在某个字母表基础上的语言集，不能与图灵机集形成一一对应关系。由于一台图灵机最多只能识别一种语言（有些图灵机无法识别任何语言），所以我们断定有些语言不能被任何一台图灵机识别，毕竟语言的数量比图灵机多。换一种不那么抽象的说法，就是所有程序的集合是可数的，而所有问题的集合是不可数的。因此，问题比程序多。这意味着有些问题是任何程序都无法解决的。

这个推理过程表明，有无数种语言无法被任何图灵机识别。但是，它并没有为我们提供关于这种语言的一个具体例子。为了让事情看起来不那么抽象，我将描述一个特别的不可判定问题，并提供几个看似简单但也不可判定的例子。

考虑这样一个问题（叫作停机问题）：已知一台图灵机M和一个字符串w，M能接受w吗？我们将会看到，任何图灵机都无法判定这个问题。假设这个问题有一个判定器（我们称之为H），判定器的输入是关于M的描述（<M>）和字符串w，即<M>：w。<M>：w只表示关于M的描述<M>和字符串w的连接。当且仅当M接受w时，判定器H才会接受输入（即进入接受状态并停机）。如果M不接受w，那么判定器H进入拒绝状态并停机。显然，H是一台图灵机，它可以判定另一台图灵机M是否接受给定的输入w。

如果存在这样的判定器H，我们就可以在它的基础上稍加改造，构建出另一台按如下方式工作的图灵机Z：输入关于图灵机X的描述<X>，Z就会像H在输入为<X>：<X>（关于图灵机X的两次描述的连接）时那样运行。这样一来，如果H接受，Z就会拒绝；如果H拒绝，Z就会接受。只要调换H在停机状态下的接受和拒绝标签，就很容易在H的基础上构建出Z。注意，输入<X>：<X>并运行H，这样的操作并不少见。H期望的输入是关于一台图灵机的描述和一个字符串w。Z是H稍加改造的产物，它的输入是关于X的描述<X>和一个字符串，在本例中，这个字符串恰好与< X >相对应。换句话说，图灵Z可以判定图灵机X是否接受关于它

自己的描述作为输入。如果X接受<X>，那么Z进入拒绝状态并停机，否则Z进入接受状态并停机。

显而易见，我们可以将任意的输入加诸于Z。当输入<Z>并运行Z时，会发生什么呢？如果Z拒绝<Z>，那么Z必然处于接受状态；如果Z接受<Z>，那么Z必然处于拒绝状态。显而易见，这是自相矛盾的，所以这样的机器根本不存在。避免这种矛盾的唯一方法就是承认Z不可能存在。由于Z是在H的基础上构造而成的，因此可以断定停机问题的判定器H不存在。既然停机问题没有判定器，就说明它不可判定。换一种不太正式但更易于理解的说法，就是我们不可能构建出程序A，帮助我们判定另一个程序B在接收到某个输入时是否会停机。

由于停机问题本身就非常抽象，让我们来看另一个具体一点儿的例子，它来自数学家戴维·希尔伯特。在1900年的巴黎国际数学家大会上，希尔伯特提出了23个数学难题，其中第10个问题是：已知一个多项式形式的丢番图方程，它有多个变量和整数系数，这些变量是否有整数解，从而使多项式的值为0？例如，已知多项式为

$$xy^3 + x^5 - 4x^2y^3 + 7,$$

是否有可能找到x和y的整数解，使该多项式的值为0？希尔伯特要求找出可以通过有限步骤（比如算法），成功地判定这类方程是否有解（答案只能为“是”或“否”）的方法。在不失一般性的前提下，为了简化讨论，我们可以假设x和y的整数解是正整数。

从图灵机的角度看，这个问题必须用某种语言来表示，即字符串集。该语言中的每个字符串都表示这个问题的某个实例（在本例中是某个丢番图方程）的编码。因此，与希尔伯特的第10个问题对应的语言是一个字符串集，其中每个字符串都是该问题实例的编码。例如，我们可以判定字符串{x, 2, y, 1, 0, +, 1, 5}是多项式$x^2y^{10}+15$的编码。这种语言拥有无穷多个字符串，因为多项式的数量是无穷的。

根据对图灵机的了解，我们也许能够想出一个解决办法。我们可以从写在纸带上的多项式输入着手。图灵机将以某个特定序列的方式，例如，(0, 0)，(0, 1)，(1, 0)，(0, 2)，(1, 1)，(2, 0)，…，尝试x和y值的所有可能组合。对于每个组合，图灵机都会计算多项式的值。如果值为0，图灵机就会停止运行，否则它会继续尝试下一个组合。运用这种方法，一旦找到解图灵机就会停机，但如果该问题无解，图灵机就会一直运行下去。有人可能认为，稍加改进即可轻松解决这个小问题。但是，通过马丁·戴维斯、尤里·马季亚谢维奇、希拉里·普特南和朱莉娅·罗宾逊等人的研究，我们现在知道这个问题是不可判定的。这意味着没有图灵机可以识别这种语言中的字符串，也就是说，肯定会停机（当字符串对应于一个有整数解的丢番图方程时就会进入接受状态，而当字符串对应于一个没有整数解的方程时则会进入拒绝状态）的图灵机是不存在的。必须明确的是，对于某些特殊方程，我们有可能确定它们是否有解。但一般性问题是不可判定的，我们不可能确定它们是否有整数解。

有人可能认为，只有非常抽象的逻辑或数学问题才是不可判定的。事实并非如此。接下来，我再介绍一个看似更简单但也不可判定的问题——波斯特对应问题，以此结束本节的讨论。

假设给你一副多米诺骨牌，用来搭积木。一块骨牌的上半部和下半部分别刻有一个特定的字符串。例如，一块骨牌的上半部刻着abc，下半部刻着b。波斯特对应问题假设有这样一副多米诺骨牌，而且每张骨牌的拷贝数量足够多，需要我们确定是否存在一个多米诺骨牌序列，当它们并排放置时，上半部和下半部可以拼写出相同的字符串。比如，假设给定的骨牌集合是{abc/b, f/cfa, c/cc, e/ea}。是否存在一个骨牌序列（可以重复使用同一张骨牌），它们的上半部和下半部可以拼写出相同的字符串？就这幅骨牌而言，答案是肯定的。如图4－6所示，序列{e/ea, abc/b, f/cfa, abc/b, c/cc}的上半部和下半部都能拼写出的字符串是“eabcfabcc”。

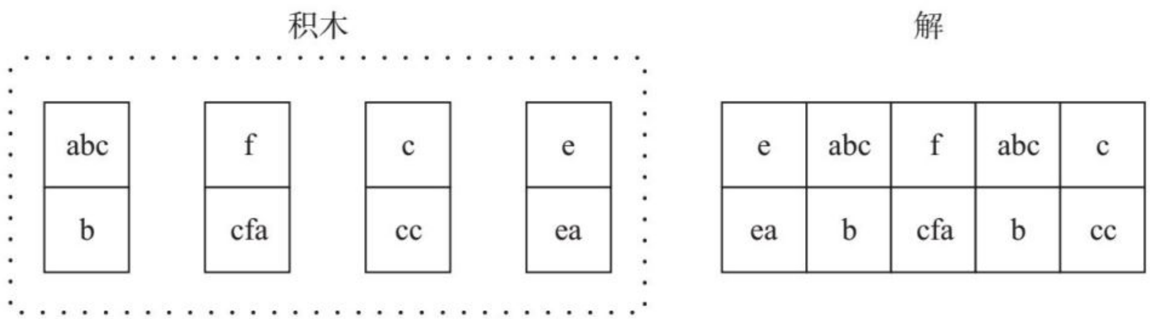


图4－6 已知不可判定的波斯特对应问题示意图

事实证明，与刚刚讨论的这个简单谜题相对应的问题也是不可判定的，这意味着没有任何算法能保证有效。如果真的存在一种可以解决这个问题的算法，那么它也可以用来解决停机问题，但矛盾的是，停机问题表明根本不存在这样的算法。注意，在某些情况下，给出肯定或否定的明确回答是有可能的，但要设计出关于这个问题的一种通用算法，确保总能给出“是”或“否”的准确答案并让图灵机停机，是不可能的。

算法和复杂性

我们已经知道有些问题是不可判定的，也就是说，没有一种算法可以保证在所有情况下都能找到这些问题的答案。而其他可以判定的问题——我们在日常生活中遇到的大多数问题都属于这一类——必然可以找到算法来解答。通常，算法被设计成抽象的操作序列，没有明确提及底层计算支持（可以是存储程序计算机、图灵机或其他计算模型）。这些算法的效率或高或低，花费的时间或多或少，但最终它们会停止运行，并给出一个答案。

我已经讨论了算法及其在电子电路设计、人工智能等领域的应用。算法设计是一个巨大而复杂的领域，也是许多技术进步的根源。先进的信号处理算法使手机、卫星通信和安全的互联网通信成为可能。计算机辅助设计算法有助于工程师设计计算机、汽车、轮船和飞机。优化算法与我们购买和使用的几乎所有东西的物流和配送有关。在许许多多的情况下，无论是我们动脑规划去最近咖啡店的路线时，还是在接电话时，我们都在无意识地使用算法。

算法是一系列步骤，给定一个输入，就会得到某个特定的结果。为了直观地说明这个问题，我将举几个在城市、地图和道路距离等特定领域中常用的算法实例。从这些算法的角度看，城市、道路和距离的地图可以用图来表示。图是一种数学抽象，可用于表现许多问题和领域。图包含一系列节点和边，边连接成对的节点。节点和边都可以有权重；边可以有向，也可以无向，这取决于问题的性质。连接给定节点的边数被称为该节点的度。在有向图中，输入顶点数被称为入度，输出顶点数被称为出度。

假设图4－7表示法国多座城市之间的道路和距离。现在，考虑一个常见问题，即在地图上找出某两座城市之间的最短距离。给定任意两座城市，算法都可以找出它们之间的最短路径。例如，如果某个人想从巴黎去里昂，最短路线就是途径第戎，总距离为437公里（从巴黎到第戎为263公里，从第戎到里昂为174公里）。找出任意两座城市之间的最短路径比较容易，有若干算法都可以高效地完成这个计算任务。1956年，艾兹格·迪科斯彻提出了第一个符合要求的算法，它可以找出从一个起始节点到任意节点的最短路径。随着图不断变大，问题变得越来越有挑战性，算法的运行时间也不断增加；但是，这些增长都是渐进的，即使是非常大的图，算法也可以搞定。

关于这幅图的另一个问题也很容易回答：如何让连接所有城市的道路总长度最短？这个问题被称作最小生成树，它有很多实际应用，比如大脑网络研究。后来，用于构建摩拉维亚的高效供电网络的方法，成为解决这个问题的第一个算法。

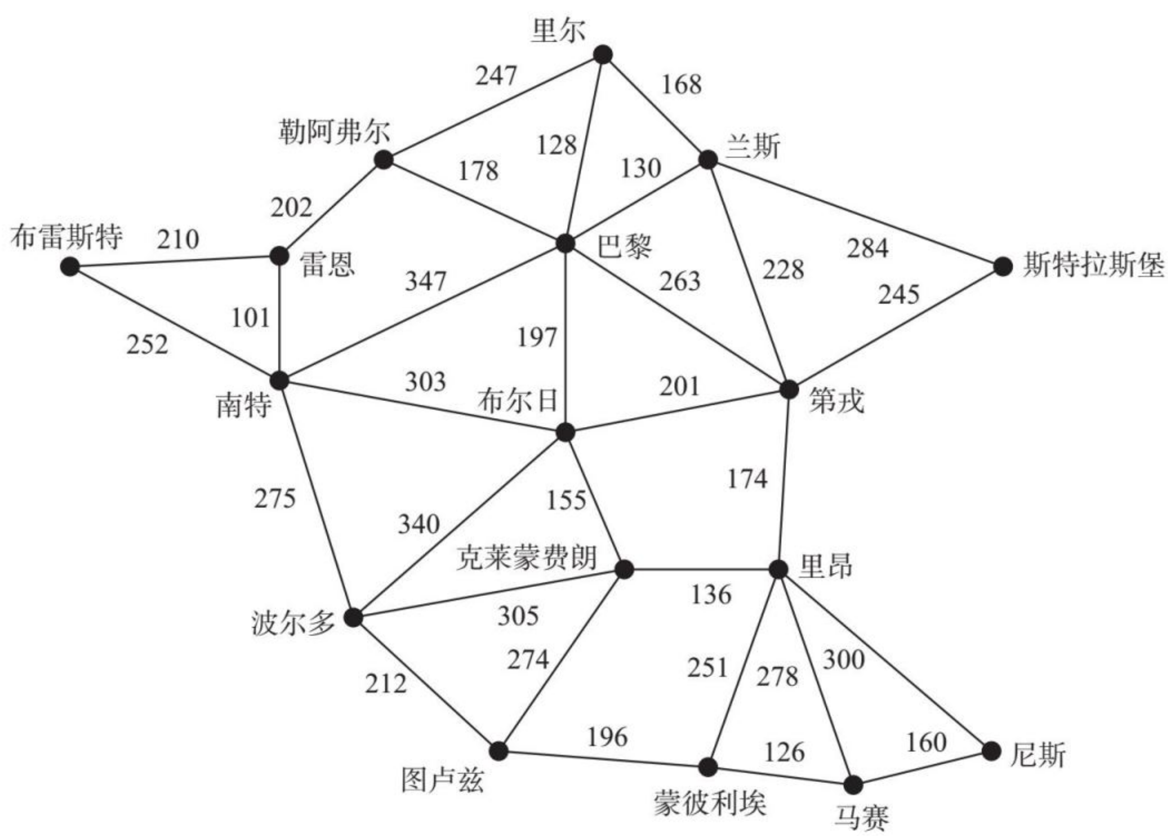


图4－7 法国城市间距图

最小生成树可以通过几种非常有效的算法找到，其中普里姆算法可能是最简单的一种。它依次选择尚未被选择且选择后不会构成环路的最短道路，并将其纳入生成树。当所有城市都被连接起来后，算法就会停止运行。将该算法应用于图4－7，就会产成图4－8所示的最小生成树。

其他更具挑战性的问题也可以用这幅图来阐述。假设你是一名推销员，需要走访每座城市恰好一次，然后返回起点城市。这就是哈密顿圈问题，它以威廉·罗文·哈密顿的名字命名。哈密顿发明了一种游戏（被称为哈密顿周游世界问题），要求玩家在十二面体的边图中找出一个圈。随着图的尺寸增大，在图中找出哈密顿圈的复杂程度迅速增加。如果你想沿最短的路线，以尽可能快的速度走访所有城市，就需要解决所谓的旅行商问题。同样，当只有几座城市时，问题很简单，但当有很多座城市时，问题的难度就会大大增加。

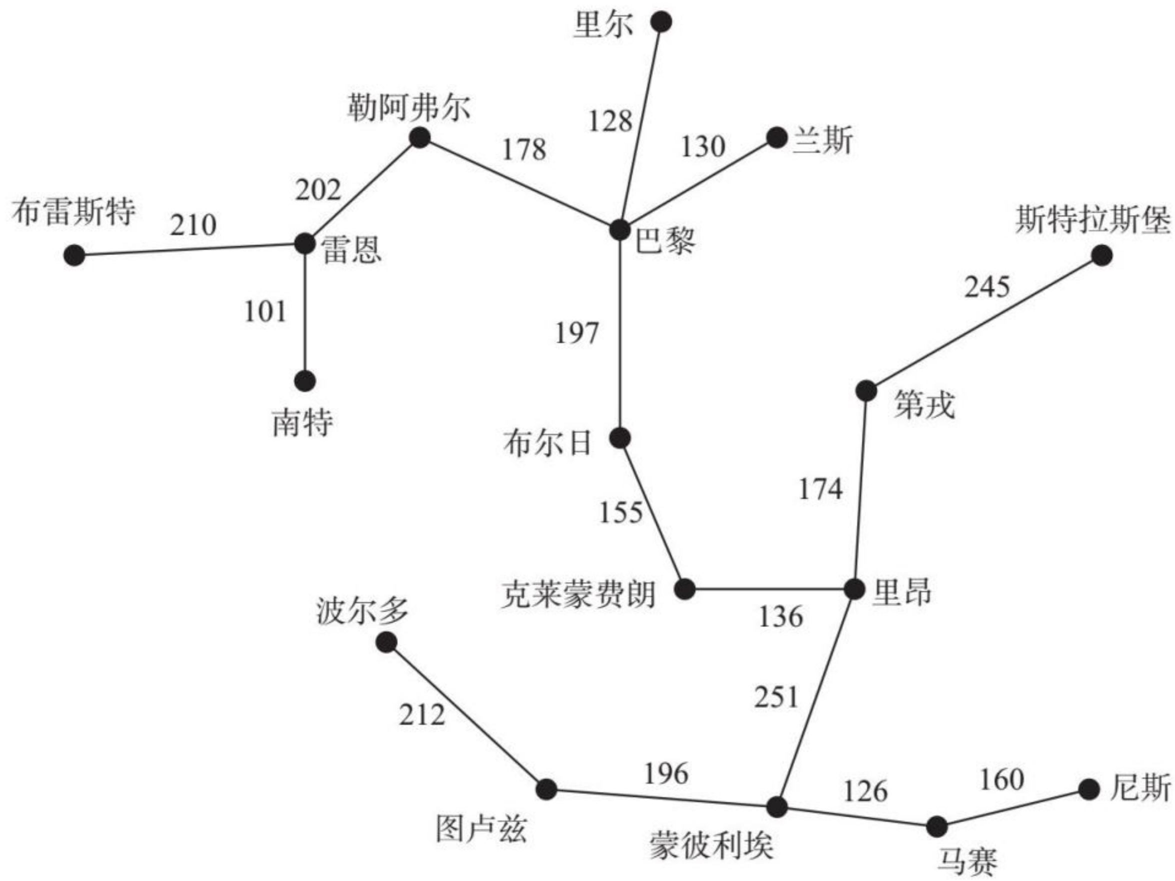
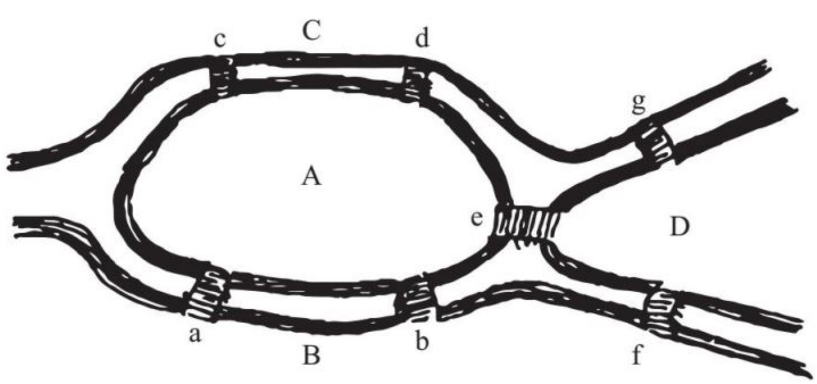


图4－8 图4－7所示法国城市间距图的最小生成树

还有一个可以用图阐述的简单问题是欧拉路径问题，这个名字可能会让我们联想到1736年莱昂哈德·欧拉解决的那个著名的哥尼斯堡七桥问题：哥尼斯堡市有7座桥，连接着普雷格尔河的两岸和两座岛屿（图4－9），需要找出一条穿过这座城市的路，每座桥都要走一次，而且只能走一次。这个问题可以用一幅图来表示，图中的4个节点分别表示两座岛和河的两岸，边表示桥。（本例中的图叫作多重图，因为在相同的两个节点之间有多条边。）经过图中每条边正好一次的路径被称为欧拉路径，返回起点的欧拉路径被称为欧拉圈。通过表明当且仅当奇数度顶点的个数恰好为0或2时，图中才会存在欧拉路径，欧拉证明不可能找到一条穿过哥尼斯堡市且每座桥恰好走一次的路径。如果奇数度顶点的个数为0，那么所有欧拉路径也都是欧拉圈。

设计算法就是设计一系列步骤的技巧，如果这样做了，就能以最高效的方式找到正确答案。从根本上说，算法的效率就是它们在计算机上运行时花费的时间。当然，运行一个算法需要花费的实际时间取决于我们使用的计算机。但是，如果待解决的是一个很大的问题，而算法又非常低效，即使是速度最快的计算机，也需要花费很长的时间；反之，如果算法非常高效，即使是速度较慢的计算机，也可以解决一个合理大小的问题。这是因为当算法的计算复杂度高时，它的运行时间就会随着问题的增大而快速增加。算法的计算复杂性是通过它们的规模随问题尺寸的变化方式来衡量的。问题的尺寸通常就是描述该问题所需的输入量。在这个例子中，图的大小与问题描述中的节点（位置）和边（道路或桥）的数量有关。严格来说，我们还应该考虑关于每个节点和每条边的描述的长度，但在很多情况下，我们可以假设这个因素不太重要，因为只需要很少且数量恒定的字节就可以描述每个节点和每条边。例如，如果问题的尺寸为N，算法的执行时间随问题的尺寸呈线性增长，我们就说该算法的复杂性与N同阶。但是，算法的运行时间也有可能随问题尺寸的平方增加。在这种情况下，算法的复杂性与N^2同阶；当问题的尺寸增加一倍时，算法的运行时间就会增加为原来的4倍。不仅如此，算法运行时间的增长速度甚至会更快。例如，当算法的复杂性为2^N阶时，即使问题的尺寸只增加1，算法的运行时间也会增加为原来的两倍。



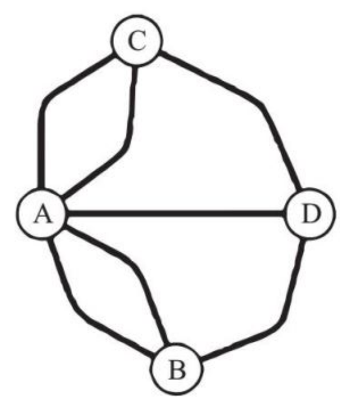


图4－9 欧拉手绘图和哥尼斯堡七桥问题示意图

就执行时间的增长速度而言，具有多项式复杂性（比如N^2）的算法和具有指数复杂性（比如2^N）的算法，两者之间的差别很大。一方面，即使使用的计算机速度很快，而且处理的只是小问题，具有指数复杂性的算法也会变得极其耗时。另一方面，具有多项式复杂性的算法通常不需要花费过多的时间，就可以解决很大的问题。

假设有一台每秒可执行10亿（10^9）次操作的计算机，以及三种不同的算法，它们解决大小为N的问题分别需要N、N^2和2^N次操作。不难看出，当问题的大小为1 000时，需要N次操作的算法将花费1微秒的计算时间。尽管如此，我们还是有必要计算一下这三种算法的执行时间是如何随N增加的。表4－5给出了这台假想计算机运用不同复杂性的算法解决这个问题所需的时间。如果运行时间大于宇宙的已知年龄（大约140亿年），表中就会出现“∞”符号。我们可以清楚地看出，即使是处理中等大小的问题，具有指数复杂性的算法也需要花很长时间。这是我们熟悉的指数函数性质产生的结果：即使底数很小，指数函数的增长也非常快。

表4–5 不同复杂性的算法的执行时间。其中，∞表示比宇宙年龄还长的时间

问题大小	算法复杂性		
	N	N^2	2^N
10	10纳秒	100纳秒	1微秒
20	20纳秒	400纳秒	1毫秒
50	50纳秒	2.5微秒	13天
100	100纳秒	10微秒	∞
200	200纳秒	40微秒	∞
500	500纳秒	250微秒	∞
1000	1微秒	1毫秒	∞
1 000 000	1毫秒	17分钟	∞
1 000 000 000	1秒	32年	∞

相反，具有多项式复杂性的算法可以解决很大的问题，虽然当问题很大且多项式的次数大于1时，它们可能也需要花很长时间。因此，多项式时间算法通常被称为有效算法，尽管在某些情况下，它们仍然需要运行很长时间才会结束。能被多项式时间算法解决的问题叫作易解问题，否则就是难解问题。一般来说，这可以将问题清楚地分成两大类，一类问题有多项式时间算法，另一类问题则没有已知的多项式时间算法。

两座城市之间的最短路径问题和旅行商问题分属两个不同的复杂性类别。事实上，根据目前已知的情况，有若干多项式时间算法可以解决最短路径问题，但只有指数时间算法可以解决旅行商问题。欧拉路径问题也可以在多项式时间内得到解决，而且它的算法只随图的大小呈线性增长。

有人可能认为，通过进一步研究，我们有可能设计出一个可以解决旅行商问题的多项式时间算法。但是，这个想法几乎不可能实现，因为在某种意义上，旅行商问题所属的那一大类问题都具有相同的复杂性。这类问题被称为NP，已知其中有很多问题都没有有效解。NP的

意思是非确定性多项式时间，之所以叫这个名字，是因为非确定型图灵机可以在多项式时间内解决这类判定问题。非确定型图灵机指的是每时每刻都可以采取许多行动，并（非确定性地）将状态更改为许多不同状态的机器。判定问题指的是可以表述成只接受“是”或“否”这两种答案的问题。

与前文中描述的确定型图灵机相比，非确定型图灵机更奇怪，也更不实用。非确定型图灵机只是一种概念装置，不可能被建造出来，因为每时每刻它都可以转换成若干状态，同时写出若干符号——可以说它类似于平行宇宙。随着时间的推移，非确定型图灵机使用的格局数量呈指数增长，因为每一个新动作，加上之前的动作，都会在计算树上形成新分支。人们相信，非确定型图灵机的速度将远超确定型图灵机。不过，我们很快就会看到，这个问题仍然悬而未决。

就当前的讨论而言，只要知道NP类问题与答案可以在多项式时间内得到有效证实的判定问题完全一致，就足够了。我们之所以能用这种方式定义NP类问题，是因为非确定型图灵机是通过“猜”答案的方式来解决这些问题的，即利用它的非确定性能力，在纸带上同时写下多个答案，然后检查这些答案是否正确。如果答案正确，图灵机就会进入接受状态并停机。如果没有找到答案，图灵机就不会进入接受状态并停机，该次计算也会被忽略。

有些NP类问题特别难，原因在于，如果其中一个问题存在有效解（在多项式时间内可解），那么其他所有问题也必然存在有效解。这些问题被称为NP困难问题，旅行商问题就属于这类问题。既是NP困难问题又是NP类的判定问题，被称为NP完全问题。

上文阐述的旅行商问题不是一个判定问题，因为它要求找出最短路径。但是，我们可以把它重新表述成一个相关的判定问题：在给定的图中，是否存在长度小于给定数字的路径？判定问题的答案可能很难找到（本例就是这样），但验证答案是否正确却很容易：对于一个

给定的答案（依次经过图中列举的城市），只需要将这条路径上的各座城市之间的距离相加，就可以验证其总长度是否小于给定数字。

P类问题是NP问题的一个子集，它包含（确定型）图灵机可以在多项式时间内解决的所有判定问题。与最短路径问题相关的判定问题就属于P类：城市A和城市B之间是否存在一条长度比给定数字短的路径？

史蒂芬·库克的研究表明，如果一个NP完全问题有一个有效算法，我们就可以利用该算法为NP类中的每一个问题，包括所有NP完全问题，设计出一个有效算法。理查德·卡普证明了另外21个问题是NP完全问题之后，已知的NP完全问题的名单不断增长。有趣的是，库克和卡普在从事上述研究工作期间都在加州大学伯克利分校任教，但库克在1970年被拒绝授予终身教职，之后仅过了一年，他就发表了他的那篇开创性论文。用理查德·卡普的话说，“我们未能说服数学系授予他终身教职，这让我们永远蒙羞。”

没有人真正知道是否存在一种可以有效解决NP完全问题的算法，但大多数科学家都认为这种算法不存在。如果有人发现了（任意）一个NP完全问题的多项式时间算法，该多项式时间算法经修改之后，就可以有效地（即在多项式时间内）解决NP类中的任何问题，由此证明NP实际上等于P。

P等于NP还是不同于NP，这是计算和数学领域中最重要的悬而未决的问题之一（甚至可以去掉“之一”两个字），库克在1971年的论文中也提到了这一点。事实上，它是克莱数学研究所在2000年提出的7个“千禧年大奖难题”之一。只要正确求解出其中任何一个问题，就能获得100万美元的奖金。

尽管人们普遍认为P不同于NP，但大多数科学家质疑这个问题在短时间内不会有确切的答案。如果NP中的某些问题不属于P（我认为这很有可能），那么这些问题将很难准确有效地解决。

确定合数的因数，就是一个已知的属于NP但不一定属于P的问题。合数是可以写成两个或两个以上的较小整数之乘积的整数。给定两个数，通过标准的长除法，可以很容易地验证第一个数是否是第二个数的因数，计算机能够非常迅速地做到这一点。

马宁德拉·阿格拉沃、尼拉杰·卡亚和尼丁比·萨克塞纳已经证明，在多项式时间内可以确定一个数是不是质数。但是，如果不知道一个非常大的整数有哪些因数，要找出这些因数就会非常困难。在这种情况下，问题描述的大小就是该数字的位数。在求解这个问题方面，所有已知算法需要花费的时间都长于多项式时间，因此分解一个位数很多的数，不仅难度非常大，而且十分耗时。有人可能认为这是一个没有现实意义的问题，但事实上，互联网通信的安全性在很大程度上取决于一个特定的密码体制——RSA算法，它依据的就是大合数的因式分解难度大的事实。如果证实P等于NP，RSA算法的安全性就有可能遭到破坏，并被目前未知的某种方法取代。

其他的计算范式——不一定等同于图灵机——可能会移动易解问题和难解问题之间的界线。量子计算是一种新的实验性计算方法，它利用量子态叠加原理并行尝试大量解，可以有效地解决因式分解问题。但是，能够解决该问题的量子计算机目前还不存在，在未来几十年内也不太可能出现。实质上是概念验证原型机的小型量子计算机每次只能处理少量比特，不能用来解决这个问题和其他所有大尺寸问题。包括我在内的许多人都认为，实用的量子计算机可能永远也不会出现。但是，鉴于本书第2章对技术进化的讨论，我们只能拭目以待，看看能否以远比图灵机强大的机器为基础，建立起某种超级图灵计算范式。

尽管目前尚未找到有效算法的NP完全问题有很多，但诸多领域都已经开发出有效算法，而且我们中的大多数人每天都在使用它们。这些算法的用途广泛：在计算机网络上发送声音和图像，将我们的想法写入计算机磁盘，在万维网上搜索信息，模拟电子电路、飞机和汽车

的行为，预测天气，以及执行许多其他任务。手机、电视和计算机都离不开算法，它们通过中央处理器执行这些算法。在某些情况下，比如旅行商问题，我们不可能找到最优解，但可以找到所谓的近似解，即接近于最优解的解。

事实上，算法无处不在，如果没有它们，现代生活就会截然不同。在计算领域得到普遍认可的丘奇–图灵论题指出，任何可计算的问题都可以通过计算机中运行的算法来解决。这可能也适用于人类擅长而计算机不擅长的问题，比如识别人脸、理解口语语句或驾驶汽车。计算机确实不擅长做这些事，但这并不意味人类永远能保持出类拔萃。相反，这可能意味着我们尚未找到适用于这些事的算法。此外，它可能还意味着人类所做的计算与图灵机进行的计算不是一回事。

丘奇-图灵论题

接下来，我们讨论一个更复杂的问题：什么是可有效计算的东西？这个问题有着悠久的历史，许多杰出的思想家和数学家都研究过它。阿兰·图灵、阿隆佐·丘奇和库尔特·哥德尔在推动这项研究并得出答案方面，发挥了非常重要的作用。丘奇－图灵论题源于图灵和丘奇于1936年发表的论文，尽管无法用数学方法加以证明，但它得到了科学界的广泛认可。丘奇－图灵论题指出，所有可有效计算的函数都可以用图灵机计算。也就是说，除了有效性上的差异，任何可计算的东西都可以用通用图灵机来计算。除了借助已知与图灵模型等价的模型之外，没有任何方法可以正式地定义什么是“可有效计算”，因此丘奇－图灵论题也被视为“可有效计算性”的一种定义。

准确地说，可计算性的概念只适用于对整数进行的计算。现存的所有计算机都是在整数的基础上工作的，因为整数可被组合起来非常近似地表示实数和物理实在的其他表达形式。这就提出了一个问题：使用物理量的物理系统会不会比图灵机更强大，可以执行图灵机无法执行的计算呢？

物理版本的丘奇－图灵论题是丘奇－图灵论题的加强版，它指出所有可进行物理计算的函数都可以用图灵机计算。虽然人们一致认为丘奇－图灵论题是有理有据的，但关于物理版本的丘奇－图灵论题的有效性还没有定论。例如，我们可以证明用实数执行计算的机器比图灵机更强大。有些计算模型的确有可能比图灵机更强大，但我们还不清楚它们能否变成实物。因此，我们根本不知道是否存在比图灵机更强大的物理系统。

物理版本的丘奇－图灵论题是否正确，直接关系到本书讨论的一个重要的开放性问题：人脑能计算不可计算的函数吗？人脑的物理结构能否让它比图灵机更强大？一般而言，运用了图灵模型未考虑或未使用的一些自然规律的物理系统，会不会比图灵机更强大呢？

1961年，约翰·卢卡斯发表的一篇题为《思维、机器与哥德尔》（*Minds, Machines and Gödel*）的论文指出，哥德尔第一不完全性定理认为人类思维比所有图灵机都强大，因为任何足够强大的数学系统都不可能兼具一致性和完全性。哥德尔不完全性定理建立在这样一个事实的基础之上：在任何给定的形式系统S中，陈述“G在系统S中无法被证明”的哥德尔语句G可以用语言S来表述。这样一来，要么这个语句是真实的但无法被证明，从而表明S具有不完全性；要么这个语句可以在S中得到证明，从而表明S具有不一致性（因为G指出G无法在S中被证明）。

卢卡斯认为，图灵机除了处理符号之外，其他什么都做不了，因此它相当于某个形式系统S。如果我们构建出S的哥德尔语句，它就无法在系统S中得到证明，除非S具有不一致性。但是，具有一致性的人类可以理解S，并“看出”这个语句是正确的，即使它无法在S中得到证明。因此，人类不可能等同于任何形式系统S，人类思维比所有图灵机都强大。

卢卡斯的理论有若干不足之处，因此人们提出了许多反对意见。（其中一个反对意见认为，人类不一定具有一致性。）此外，要“看出”一个语句在S中是有效的，人类必须能够彻底理解S，而在一般情况下，我们根本无法确定人类是否具备这种能力。因此，卢卡斯认为人类思维比图灵机更强大的观点，并没有得到太多的关注。

1989年，罗杰·彭罗斯在他的《皇帝新脑》一书中，重新审视了卢卡斯的观点。彭罗斯运用的论证方法与卢卡斯基本相同，他认为人类的意识不是算法，因此人脑不是图灵机的等价物。他猜想量子力学在人类意识中起着至关重要的作用，结构上的量子波函数坍缩使人脑

比图灵机更强大。彭罗斯指出，人脑可以利用量子物理系统的性质，去执行任何图灵机都无法仿真的计算。

到目前为止，我们还没有找到任何证据证明人脑可以利用量子效应执行图灵机无法完成的计算。我个人坚信，人脑实际上是与图灵机等价的。持有这种立场的人不只有我一个。科学界的绝大多数人都以明确或含蓄的方式表明，人脑是与图灵机等价的，只不过技术与知识方面的限制导致我们无法在计算机中重建大脑的工作机制。

在后文中我们将更深入地研究大脑的工作方式。接下来的几章将会告诉我们，大脑实际上是一台极其复杂的模拟计算机，数字计算机可以再现大脑执行的计算，而且可以达到我们想要的精确度。只要对人脑及其接收到的输入进行足够精确的描述，数字计算机原则上就能仿真人脑，而且其精确度之高足以使数字计算机给出的结果与人脑得出的结果难以区分。如果这对大脑来说如此，那么对于更简单的物理系统，特别是个体细胞和身体亦如此。认为智能行为可能是计算机的操作结果的观点由来已久，其主要原因在于，人脑和计算机执行的计算在某种意义上是等价的。但是，即使人脑比计算机更强大，人类也仍然有可能设计出智能机器。50多年来，设计智能机器一直是众多研究人员孜孜以求的目标。人类探索智能机器的历史是一部充斥着诺言、成功和失败的漫长历史，我们将在下一章讨论这方面的内容。

第5章

智能机器的探索之旅



计算机可以拥有智能吗？在计算机中运行的程序能像人类一样聪明地行动吗？这些都不是简单的问题，几百年来科学家和思想家苦寻答案而不得。荷马的《伊利亚特》最早提及会思考的机器（赫菲斯托斯的自动轮式三脚架），其他文学作品（包括玛丽·雪莱于1818年出版的小说《科学怪人》）中也描写了类似的机器，但直到多年以后人们才开始认真思考机器能否表现出智能行为的问题。阿达·洛芙莱斯在为梅纳布雷亚的《查尔斯·巴贝奇发明的分析机概述》做注释时，似乎对这个问题给出了否定的回答，她指出分析机只会执行指令，而不会发明任何东西。由于创造力通常被视为智能的标志之一，因此这似乎表明阿达·洛芙莱斯认为计算机不可能拥有智能。

对于这个问题，后来有更多的研究人员持更加乐观，有时甚至是过于乐观的态度——许多人预测智能机器将在2000年出现。

人工智能

现代人工智能（AI）研究始于20世纪50年代中期。1956年夏天，在达特茅斯学院召开的一次会议使人们对这个领域产生了极大的热情。许多与会者后来都成为该领域的领军人物，其中包括马文·明斯基、赫伯特·西蒙、约翰·麦卡锡和艾伦·纽厄尔。一些权威大学和研究机构，包括麻省理工学院、卡内基－梅隆大学、加州大学伯克利分校和斯坦福大学，先后建立了人工智能实验室。许多研究人员开始编写程序，尝试解决之前被认为需要用到智能的诸多问题。纽厄尔和西蒙编写的“逻辑理论家”程序可以证明数学定理，包括怀特海和伯特兰·罗素在《数学原理》中列出的一些数学定理。最终，“逻辑理论家”找到了怀特海和罗素在这部杰作中列出的几十条定理的证明方法，有些比当时已知的证明方法更简洁。亚瑟·塞缪尔编写的“跳棋”程序通过α－β搜索——一种搜索游戏树的方法——化身为一名相当厉害的跳棋手，其水平之高足以战胜它的程序设计员。最终，人工智能系统变得精通下棋、规划行动、调度，以及执行其他复杂的任务。

然而，智能是一个比人们曾经以为的更难以捉摸的概念。人们通常认为，解决上述任何问题都离不开智能，但我们根本不知道计算机用于解决这些问题的技术，是否赋予了它们通用类人智能。事实上，这些问题是用专门的方法来处理的，而且总的来说这些方法与人类使用的方法大不相同。例如，下国际象棋的计算机对未来的可能局面进行非常广泛的搜索，利用其极快的速度每秒评估几百万个局面，而人类冠军几乎不可能采取这种策略。计算机在进行语音识别和人脸识别时也会使用类似的专门技术。

早期的研究人员试图通过建立符号操作系统来解决人工智能的问题。他们的想法是，通过构建某种程序，按部就班地模仿人类的智能行为。他们研究并最终解决了人类日常处理的某些简单问题，比如演绎、推理、规划和调度。但是，这些任务也给研究人员带来了意想不到的困难，因为它们大多难以表述，还有一些难以计算。事实上，我们日常生活中解决的许多问题都是难解问题，因为它们都属于NP困难问题。例如，规划一系列相互依赖的行动，使世界从当前状态变成预期状态，这就是一个难以计算的问题，也是一个已知的NP困难问题。但是，人类每天都会解决NP困难问题，而且无须花费太多时间。我们不能把它视为人类大脑比计算机更强大的证据，而只能把它看作启发式方法（在大多数情况下都有效，而且能找出近似解）经过数百万年的进化，已经发展并融入了人类推理过程的证据。

尽管如此，通过开发尖端的技术，人工智能研究人员成功地解决了许多重要问题，这些解决方案现在也被应用于诸多领域。例如，列车和航班时刻表通常就是由基于人工智能的系统设计的，许多企业还使用人工智能研究人员开发的技术进行数据挖掘。语音识别系统起源于人工智能领域，目前得到了广泛应用。事实上，人工智能技术已经无处不在，几乎每天都有新的应用程序问世。

但是，设计一台公认的智能机器，仍然是人工智能研究人员无法实现的一个目标。一个重要的原因在于，智能是一个让人捉摸不透的概念。在国际象棋程序达到目前的水平之前，人们普遍认为要达到国际象棋锦标赛的水平，就需要强人工智能，即类人智能。但是，在为下棋设计特定的程序（算法）时，人们逐渐发现，无须强人工智能，使用蛮力搜索算法和尖端的局面评估启发式算法就可以达到目的。在研究诸如规划、语音理解、人脸识别和定理证明等许多其他问题的过程中，我们对人工智能的理解也发生了类似的变化。在某种程度上，我们似乎认为强人工智能难以实现的原因在于，人工智能研究人员尚未找到恰当的算法。

强人工智能难以实现，可能还有一个更重要的原因。我们之所以认为人类的行为方式是智能的，是因为人类能够把来自许多不同领域的知识和经验相互联系在一起。即使是像理解言语这种简单的行为，也需要在人的头脑中建立一个完整的世界模型。一个程序只有存储了一个完整的与人类所用模型类似的世界模型，我们才会认为它具有类人智能。建立这样的模型是一项非常复杂的工程。人类通过进化在头脑中构建了世界模型，并持续利用这些模型来消除感知的歧义，预测行为的结果，以及制订计划。然而，我们并不知道如何以明确的方式构建这样的模型。这也解释了为什么这些早期的探索人工智能的方法，只能产生非常有限的结果。

我们如何判定一个程序是否拥有智能呢？如果一台机器能够识别和合成语音，也能技术精湛地下国际象棋，难道我们不应该认为它拥有智能吗？我们如何将智能机器与只能运行某些特定算法及执行给定任务的机器区分开来呢？如果一个人能用英语和我们交谈，国际象棋下得也相当不错，即使我们对他在其他领域的能力存疑，我们肯定也会认为他是智能的。我们是不是过于以人类为中心，而对机器的要求又太高呢？

除了对计算理论做出了重大贡献以外，阿兰·图灵还提出了如何判断机器是否拥有智能的问题。在他最初的一项分析中，他假设机器最终可以达到相当不错的下棋水平。然后，他提出了一个问题：如果人类观察者只能看到棋盘上棋子的移动，那么他能判断出哪一步是机器下的，哪一步是那个糟糕的人类玩家下的吗？

最终，图灵的想法演变为图灵测试。1950年，在他的开创性论文《计算机器与智能》中，图灵提出要解决“机器会思考吗？”的问题。他没有尝试定义“智能”、“机器”等难以捉摸的概念，而是将这个问题更改为“机器能做人类（作为会思考的实体）做的事吗？”。

图灵测试的灵感来自于模仿游戏。模仿游戏是一种派对游戏：一名男性和一名女性分别进入两个不同的房间，然后参加派对的宾客向他们提问，他们通过打字来回答宾客的问题，最后宾客根据他们的回答判断他们分别在哪个房间里。在原始游戏中，一名玩家要想办法诱导提问者做出错误的判断，而另一名玩家则要帮助提问者做出正确的判断。图灵提出用一台机器取代那名女性玩家，而且那名男性玩家和这台机器都要想方设法让宾客相信他们是人类。后来，图灵建议由一个评判小组向人和计算机提问。如果评判小组中有相当大比例的人认为计算机是人，就代表它通过了测试。

图灵测试之所以经得起时间的考验，是因为它规避了我们对智能的定义中可能包含的极其明显的人类中心主义的偏见。至少，评判小组不会简单地根据它的外观来判断某个事物是不是拥有智能。

然而，图灵测试仍然带有强烈的人类中心主义偏见，因为它迫使计算机模仿人类的行为。出于这个原因以及其他一些原因，即使对今天最先进的人工智能程序来说，图灵测试的难度仍然非常大。要通过图灵测试，计算机必须拥有类似于人类的推理能力、记忆力、感觉和情感，因为评判小组向它提问的问题没有任何限制。最终，任何非人类的行为（比如情感缺失）都可以用来区分计算机和人类。例如，评判小组可以问程序它最久远的记忆是什么，最痛苦的经历是什么，或者它是否喜欢吃寿司，等等。

图灵预见了他的提议将面临的大多数反对意见，并将它们划分为九大类。事实上，所有反对图灵测试和人工智能可能性的观点，都可以归入其中的一个大类。这些反对意见值得我们详加讨论，其中有一些易于应对，因为它们没有科学依据，要么是形而上学的，要么完全不合理。

来自神学的反对意见认为，思维是人类不朽灵魂的产物，因此不能被机器模拟。它基于这样一个假设：人类在宇宙中是独一无二的，是唯一有灵魂的生物。

基于行为的非形式性的反对意见认为，没有任何规则可以描述人类在所有可能情况下的行为，这意味着人类的行为永远不可能被计算机模拟。在某种程度上，它等价于人类智能是非算法的观点，这使得它与下面的一个观点非常相似。

逃避现实的反对意见认为，会思考的机器将会导致极其可怕的结果，它之所以永远不会出现，原因可能是人类有意避免它的发展。但是，人类似乎还没有为规避风险而成功避开某项技术的先例。这个观点与我们将在第12章中简要介绍的超级智能的风险有关。

有两种反对意见基于同一种观点，即大脑不是图灵机的等价物，理由要么是大脑能（以某种不明确的方式）计算不可计算的函数，要么是因为脑细胞处理实值信号的固有能力让它比图灵机更强大。我们在第8章、第9章分析人脑工作原理和人脑仿真面临的问题时，将会分析这种反对意见。但必须指出的是，许多人都坚信这个观点是对的，我们也必须承认这个问题在很大程度上仍未得到解决。

有一种反对意见最早是由阿达·洛芙莱斯提出的，理由是计算机必须遵循固定的规则，因此它们不可能有独创性，而且它们的行为总会产生可预测的结果。这种反对意见忽略了现代工程师和科学家都很清楚的一个事实：即使完全是由固定规则定义的，极其复杂的系统也会产生完全不可预测的行为。

还有一种反对意见认为，人类拥有超感官知觉，它是机器无法仿真的。由于超感官知觉只在控制条件下才能观测到，因此这种反对意见无足轻重。

最后一个是来自哲学的反对意见，它可能比其他反对意见都深刻。它认为智力只能源自意识，而处理符号的机器永远无法获得意识。多年后，约翰·塞尔在中文房间思想实验中，以更加明确的方式论述了计算机永远不可能拥有意识的观点。

塞尔的思想实验始于一个特定假设：人工智能研究人员成功地建造出一台似乎可以理解中文的计算机。计算机接受汉字输入，并按照一系列固定的规则生成汉字输出。塞尔辩称，如果计算机能够胜任这项任务，成功地骗过一个会说中文的评判小组，根据图灵测试的标准，我们就可以得出这台计算机懂中文的结论。但是，塞尔认为这个结论是错的。他让一个不懂中文的读者想象自己被关在一个封闭的房间里，面前有纸、铅笔、橡皮擦，还有一本英文版的计算机编程指南。在接收到汉字后，读者可以通过查看书中的指令来处理这些汉字输入（模拟计算机程序的行为），并生成汉字输出。由于房间操作系统可以在不懂中文的情况下通过测试，因此我们推断计算机也不懂中文。

当然，塞尔的观点中隐藏着一种强烈的人类中心主义偏见。他认为，一个系统（房间、计算机指令，以及人类或数字计算机）即使一个汉字都不认识，也可以神奇地用中文对话。只当人类在某个环节以某种方式将一个盲目的计算过程转变为一个有意识的过程时，塞尔才会将其归因于理解的魔力。然而，这恰恰是图灵试图规避的偏见，在他提出的盲测中，评判小组不知道评判对象是不是人类，也就不会受到相关偏见的影响。

图灵详细说明了评判小组和玩家之间应该如何互动。问题和答案要用打字机打出来，以避免口语理解方面的困难，而口语理解并非问题的核心。得益于近年来语音识别和语音合成技术的进步，这个难题在不久的将来可能会变得无关紧要。我们可以想象，在成熟的图灵测试中，评判小组与人类或合成化身之间的实时对话，将取代图灵规定的打字机界面。

多项实验已经证明，我们必须谨慎对待图灵表述的那些条件，切忌过于简化。1966年编写的一个名叫Eliza（伊莉莎）的简单程序，使用简单的规则来模仿心理治疗师的行为，以至于很多人都以为它是一位真正的心理治疗师。1992年编写的一个名叫Parry（佩里）的程序模

仿了偏执型精神分裂症患者的行为，看完对话的精神科医生中约有一半人信以为真。多年来，许多其他程序据说都通过了图灵测试。但是，它们只是在有严苛限制条件的图灵测试中取得了成功，而且我们有足够的理由相信，计算机还要再过几十年才能通过无限制条件的图灵测试。不过，这并不意味着它永远不会发生。

1990年，在图灵的论文《计算机器与智能》发表40周年之际，萨塞克斯大学召开了一场关于图灵测试的学术讨论会，将一大批学者和研究人员聚集在一起。这次会议取得的成果之一是创立了勒布纳奖，被评委评选为最像人类的程序可获此奖。目前的比赛形式是进行标准图灵测试，包括用打字机打出答案。在2008年的比赛中，经过5分钟的对话，Elbot（艾尔伯特）程序骗过了12名评委中的3名，让他们误以为它是人类。这与图灵最初提出的骗过“相当一部分”，即30%的评委的目标非常接近。但是，5分钟的对话并没有多大意义。后来，勒布纳奖组委会规定必须进行更长时间和更严格的测试。在我创作本书期间，2013年的勒布纳奖得主——Mitsuku（水谷）聊天机器人可以在网上与人互动。如果测试它，你很快就会发现它的确是一个计算机程序，它有一个相当肤浅的世界模型，它离通过不带任何偏见的图灵测试的真正类人智能还差得很远。

在有史以来提出的所有评估机器智能的测试中，图灵测试仍然是最独立和最公正的一个，但它也有缺点，而且许多人都反对使用它。最明显的反对意见是，测试的效力取决于评判小组的老练程度。有些人可能很容易被蒙骗，但有些人对技术和所要讨论的问题更了解，不太容易受到误导。还有一个更强烈的反对意见是，图灵测试可能会仅因为机器的智能行为与人类行为不够接近而直接将其归为非智能行为。即使是亚瑟·克拉克在他1968年出版的小说《2001太空漫游》中虚构的神秘计算机HAL（哈尔）等超人类智能，也完全有可能通不过图灵测试，因为它们的思考方式与人类的思考方式大不相同。

总的来说，图灵测试对人工智能的发展来说并非不可或缺。人工智能研究人员已经解决了许多具有立等可见的实际用途的具体问题，但还未在开发具有类人行为的程序方面投入显著努力。尽管如此，由于图灵测试引出了一些哲学问题和挑战，因此它在计算和人工智能的发展史上仍然占据着重要地位。

机器学习

人工智能研究与人类处理及解决问题的方式有关。尽管我们很难准确定义人工智能领域具体包括哪些方面，但所有这些方面都以这样或那样的方式与人类表现智能行为的方式相关。人类智能赋予我们制订计划、理解语言、解释世界和控制我们的活动（四处移动和操纵物体）的能力，这些能力在很大程度上与计划制订、自然语言处理、人工视觉和机器人技术等人工智能的子领域相对应。几十年来，人们运用多种方法对这些领域中的多数问题进行了研究，但事实证明，这些问题的难度总的来说比它们乍看上去要难得多。例如，西蒙·派珀特曾把如何描述视野中的物体的问题，作为暑期项目布置给麻省理工学院的学生，但经过数千名研究人员的努力，这个问题仍未得到解决。直到最近，人们利用这个领域的新进展创建了处理这个问题的系统，才取得了一定程度的成功。

关于人工智能及其诸多子领域的著作有很多，其中包括斯图尔特·罗素与彼得·诺维格合著的《人工智能》。人工智能领域的范围非常广，即使是研究人员多年来开发的技术与理论，我也无法在这里一一介绍。

但是，人工智能的子领域之一——机器学习——对我们正在讨论的问题而言非常重要，因此它理应在本书中占据一个特殊的位置。机器学习对所有需要智能的活动来说都不可或缺，它关心的是机器从经验中学习的能力，这是系统表现出智能行为必须具备的最重要能力。关于这个主题的好书有很多，我强烈推荐大家阅读杰罗姆·弗里德曼、特里瓦·哈斯蒂和罗伯特·提布施瓦尼的《统计学习基础》，弗

拉基米尔·万普尼克的《统计学习理论》，佩德罗·多明戈斯的《终极算法》，以及汤姆·米切尔的《机器学习》。

人类有多种学习形式，包括让我们学会阅读和写作的学习，还包括让我们学会口语、识别物体甚至走路的学习（这些学习更具先天性和原始性）。机器学习就是在程序中复制这些能力和其他相关能力。如果没有研究人员发明的机器学习技术，包括自然语言处理、人工视觉、机器人技术和智能规划在内的所有重要的人工智能领域，都不会得到发展。

学习对人工智能系统的发展具有至关重要的意义，但人们对这个问题的认识并非总像现在这样清晰。早期的研究人员试图在程序中构建完整的世界模型，使这些程序能够进行对话或处理图像。符号主义人工智能（现在被称为“有效的老式人工智能”）并没有取得预期效果，因为创建人工智能系统的难度远超预期。例如，建立一个明确和完整的世界模型，现在被视为一项难度极大的任务，因为没有人能明确地描述这样一个模型，更不要说用计算机可用的形式把它写下来了。我们每个人都有自己的世界模型，这个复杂易变的多面模型与其他人的模型大不相同。此外，事实证明，让符号系统适应现实问题中总会存在的不确定性，难度也远超预期。例如，对象识别常会受到相机噪声和光照条件的影响，但我们很难将这些现象纳入明确的逻辑模型。

最终，大多数研究人员转而认为，人工智能系统必须具备内在的适应和从经验中学习的能力，这种能力最终将成功创建出所需的世界模型。这意味着，我们不必从一开始就创建一个智能系统，我们需要创建的是一个可以根据经验调整其内部结构并通过学习变得智能的系统。这个想法同样可以追溯到阿兰·图灵，他在提出图灵测试的那篇论文中指出，编程一台会学习的机器可能比编程一台完全可用的会思考的机器更容易：

为什么要尝试设计一个能模拟成年人思维的程序，而不是一个能模拟儿童思维的程序呢？只要接受合适的教育，它就会拥有一个成年人的大脑。儿童的大脑可能就像我们刚从文具店买来的笔记本，条条框框的东西很少，而空白页有很多。我们希望儿童大脑里条条框框的东西越少越好，同样地，为像儿童大脑一样的机器编程也会比较容易。

事实上，学习是人类大脑的功能，许多其他动物的大脑也会学习，但程度逊于人类。新生儿的大脑不能执行很多动作。新生儿不会走路，不会说话，他们只有基本的视觉感知和运动技能，也无法独自在野外生存。但随着时间的流逝，大脑通过结构上的调整，帮助小孩子学会走路、说话和移动，并以一种有意义的方式与世界互动，最终成长为一个自主和自足的主体。

人类如此，除人类以外的其他灵长目动物和哺乳动物，甚至是非哺乳动物，在一定程度上亦如此。新生儿的大脑还未完全成形，因此它们适应世界的方式就是赋予其主人生存所需的任何技能。人类的独特之处在于学习过程的延伸，这个过程远长于其他所有物种，并将一个没有自卫能力的新生儿转变成一个完全自主的成年人。因此，我们学习能力的延伸是将我们的大脑与其他动物的大脑区分开的关键特征。在程序中复制这种能力是人工智能研究的一个重要领域，即机器学习领域。

机器学习的核心思想比较容易解释，一些已经发展成熟并在很多领域得到成功应用的机器学习技术同样如此。简言之，学习的精髓是，根据一些特定的、具体的经验推断出一般性的规则或行为。这被称为归纳学习法，它可以采取多种形式。归纳学习法应用广泛，是无数科学文章、书籍和学位论文（包括我自己的学位论文）的主题。

简单形式的归纳学习法是指，从一组带标签的实例中学习一般性规则。在这种情况下，每个实例都可以用一组固定属性的值和一个规

定实例类别的标签进行描述。例如，考虑一个假设性的学习问题：根据经验判断某一天是否适合打网球。你记得过去有某些天不适合打网球，于是你列出了那些天的几个天气变量，并绘制了一个表（表5－1）。为了避免在天气不适合的日子里打网球带来的负面体验，你决定依据这个表推断出一条规则，帮助你判断适合打网球的天气具有哪些特点。（当然，这是一个人为设计的问题，但它足以说明我的核心观点。可以想象，通过对这张表进行分析，你可以推断出一条规则，并据此判断某一天是适合打网球的日子还是不适合打网球的日子。）

表5–1 学习实例

温度	湿度	风速	是否适合打网球?
70	95	5	是
32	80	10	否
65	80	20	否
75	85	10	是
30	35	8	否
75	35	8	是
72	35	25	否

如果实例的数量很少，可能的解决方案或规则就会有很多，但有些看似比另一些更合理。对形式最简单的机器学习而言，它们的目标就是找到合适的分类器，在应用于尚未观测到或经历过的未来实例时会取得更好的预期表现。在这种情况下，分类器就是一条或一组规则，在给定某个实例的属性值后可以生成该实例的标签。就打网球的案例而言，实例对应的是某个特定的日子，实例的属性是那一天的天气特征，目标是用“是”或“否”标记它。

实例的可视化有时可以提供些许线索。例如，只利用温度和风速这两组值将表5－1中的实例绘制成图，就可以得到图5－1，从中可以

看出适合打网球的日子似乎都集中在一个二维空间中。事实上，这种做法可以立即产生一种简单的日期分类方法：当给一个新的实例分类时，只要在表中找出属性与它最接近的那个实例，后者的标签就是我们要找的答案。例如，如果某一天的属性值（温度，湿度，风速）=（34, 80, 12），那么它与表中第二行列出的那一天非常相似，由此可知这一天不适合打网球。这种有用的算法被称为“最近邻算法”，由此产生的大量归纳方法有时被统称为“基于相似性的学习”。但是，最近邻算法存在一些难点，致使它在许多情况下并不适用。具体来说，基于相似性的方法要求实例之间有一定程度的相似性（但有时很难找到）；如果需要搜索大量数据库才能找到最相似实例的标签，那么这些方法的效率可能会很低。

我们现在知道，根据具体的例子推断一般性规则并不是一个适定问题（well-posed problem），因为它总有诸多可能的解决方案，而且缺乏可以从中选出最佳答案的通用方法。在本例中，你也许喜欢在温暖无风的日子里打网球，但其他一些特殊的天气特征组合也有可能让你体验到这项运动的乐趣。此外，过去让你体验到打网球的乐趣的东西，将来未必如此。

大卫·休谟最早指出了归纳法存在的问题，他认为我们没有理由相信未来应该与过去相似。只有确信实例的有限集中存在某种规律性，我们才会做出推断。休谟清楚地看到，我们永远无法确定我们获得的关于某个特定现象的数据，是否足以让我们预测未来。在休谟看来，即使是像“太阳每天升起”这样广为人知的现象，也不能用来预测未来，但通过演绎法得到的数学真理显然是正确的：

> “太阳明天不会升起”的命题和“太阳明天会升起”的断言，都不难理解，也不自相矛盾。因此，我们无论如何也无法证明前一个是假命题。如果能，就说明它是自相矛盾的，而人脑显然不会构想出这样的命题。

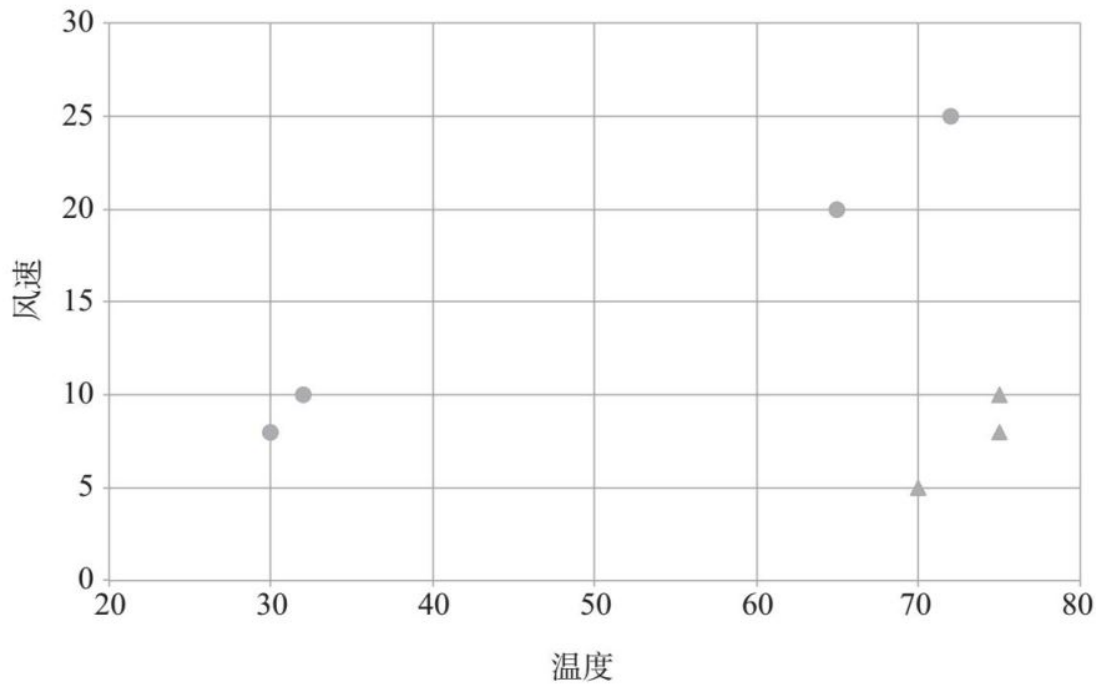


图5－1 利用温度和风速这两个属性对表5－1进行的图形化描述。三角形代表适合打网球的日子，圆形代表不适合打网球的日子

即使到了今天，休谟认为归纳过去的经验无法得到可靠结果的基本观点，在今天仍然像在18世纪那样有效。我们之所以有从经验中学习的能力，只是因为数据中存在某种规律性，而且出于某种原因，我们能够探索它。这被称为归纳偏倚，它究竟是什么的问题已经困扰了研究人员几十年，以至于诸多学派提出了各种各样的学习算法。佩德罗·多明戈斯在他2015年出版的《终极算法》一书中，系统地梳理了机器学习研究人员开发的方法和偏倚，并将它们划分成五大类，每一类都基于一个特定的理念。本章将简要介绍其中一些方法。

如果以各种学习算法解决特定领域中的所有可能问题的平均表现来衡量，或许可以证明它们之间并不存在孰优孰劣的问题。这被称为“没有免费的午餐”定理。也就是说，如果不能确定对特定的学习偏倚更偏爱，所有的学习算法就同样好。而对于现实问题，我们通常可以确定自己对某些解释的偏好程度强于其他解释。因此，我们才会从经验中学习，而且日复一日。

当然，这些问题的难度可能比上述例子大得多。事实上，它们的难度可以任意大，因为几乎所有的人类知觉经验都可以用略有不同的实例表格具体地描述为一个个归纳问题。例如，考虑识别图像中的对象这个难度极大的问题，尽管相关研究已经开展了几十年，但基本上未得到解决。从某种意义上说，这个问题是我在本章前面提到的场景识别问题的简化版。在这个更简单的问题中，我们需要找到一个分类器，去识别图片中是否有某个特定对象（比如汽车），并用“是”或“否”来回答“图片中有汽车吗？”的问题。图片本身可以是一个灰度像素阵列，像素可以取0到1之间的任意值，0意味着黑色，1意味着白色。彩色图像与此类似，但每个像素需要三个值来确定它的颜色和强度。

这个问题可以轻易地表述成典型的归纳问题，即从一组实例中习得规则。在本例中，属性就是图像中的像素。如果图像中有汽车，标签为“是”，反之则为“否”。这张表必定非常大。如果使用百万像素的图像，就会有100万个属性，每个对应表中的一列。此外，该表还有一列需要填入“是”或“否”。尽管这两个问题从表面上看很相似，它们也都可以通过实例表格直接表述成归纳学习问题，但由于各自的内在复杂性，所以它们的解决方案截然不同。

对于表5－1中的问题，我们比较容易想到一个简单的规则，它可以将接下来的每一天区分为适合打网球的日子和不适合打网球的日子。可能像温暖无风的日子适合打网球这样的简单规则，就可以很好地解决这个问题。计算机可以驾轻就熟地处理这类任务，事实上，机器学习也常被用来解决这些问题。例如，当你使用信用卡或连续从你的银行账户中取款时，银行就会利用机器学习来分析你的交易行为。它们使用的规则可能源自以前的数百万次交易。根据从你最近的一系列交易行为中观察到的模式，它们就可以确定是否应该发出警报，以及核查这些交易。计算机依据习得的一系列规则来处理这类相对简单的问题，这些规则不仅适用于训练期间使用的案例，也适用于以前从

未出现的新情况。这正是机器学习的意义所在：习得的规则可以应用于系统之前从未见过，但与它们见过的案例有一定相似性的实例。

对象识别是一个类似概念，但实际上难度要大得多。尽管人类学会了识别汽车，并且可以轻松地判断出图像中是否有汽车，但我们并不十分清楚大脑是用什么机制来执行这项任务的。据推测，在我们认识了汽车之后，我们的大脑会构建出汽车的内部表征。我们学习识别汽车的各个部分（车轮、车门、车头灯等），我们的大脑则以某种方式把这些东西组合在一起，这样一来，我们就可以识别整辆汽车了。当然，即使是车轮和车门也很难识别，但它们可能比整车更容易识别。

由于机器学习领域取得了长足发展，所以计算机程序现在已经可以识别对象了。有很多算法可用于执行归纳学习，我在这里只讨论其中几种。从模式中学习与本书的核心话题密切相关，因为人脑的可塑性主要源于相关机制。可塑性指大脑改变并适应新输入的能力，它被视为动物和人类学习的核心机制。

已知一组实例，每个实例都有一个标签，要求推导出某种规则，用于区分积极的实例和消极的实例。在处理有若干个可能的标签值的一般性实例时，也可以采用同样的方法。一个办法是，推导出能区分积极实例与消极实例的逻辑或数学公式。例如，我们可以采取“如果天气暖和无风，就是一个适合打网球的好日子”形式的简单规则，或者采取以这些简单原始规则的逻辑组合表示的更复杂规则。

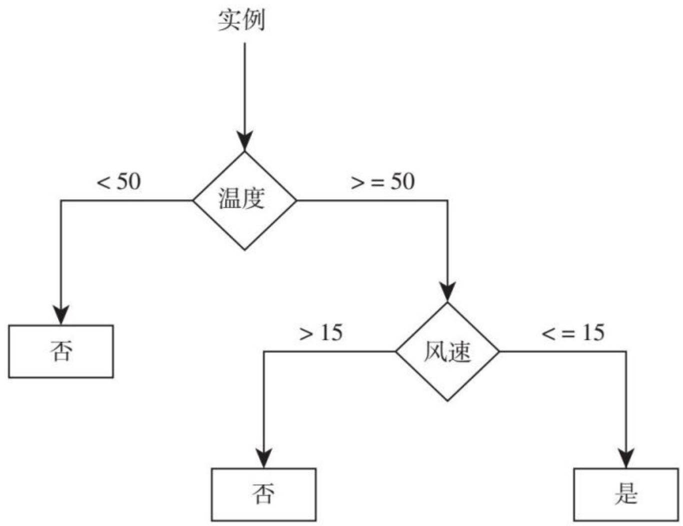


图5－2 网球问题的决策树

决策树推理是一类非常成功的学习算法，它将规则表述为一棵“树”，然后运用它对之前从未见过的实例进行分类。图5－2展示的是一个可用于分类网球问题实例的决策树。

决策树学习者试图推断出最优的可能树形，他们通常不喜欢比较复杂的树形，而更偏爱简单的树形；他们的理由是，适用于可用数据的简单解释，其预测效果强于复杂解释（这个观点最早是由奥卡姆的威廉提出来的，通常被称为“奥卡姆剃刀”）。“如无必要，勿增实体”常被解释为支持简单理论而非复杂理论，据约翰·庞奇考证，这个表述源自奥卡姆。

奥卡姆剃刀原理背后的简单直观的理念引发了许多复杂的讨论，比如，“简单”指什么，为什么简单的解释会有更强的预测效果。对于这种喜爱简单性的偏倚，哲学家、数学家和计算机科学家进行了广泛的研究和辩论。可以断定的是，简单的规则通常具有更强的预测效果，尽管“简单”的定义需要用非常明确的数学术语来阐述。

为了得到简单的解释，许多决策树学习者根据属性中包含的与标签相关的信息量多少，贪婪地选择待测试的属性。他们优先选择信息量大的属性（也就是与标签高度相关的属性），从而形成了对更小和更简单的决策树的偏好。

基于一系列测试的决策树，以一种易于理解的结构化方法表示决策程序，可以高效地应用于多个领域。因此，它们在医疗诊断、财务分析，以及许多其他科学和工业领域中得到了广泛应用。但是，就许多需要进行归纳推理的问题而言，尤其是当属性数量众多且为实数值，以及类别的界线复杂且高维时，决策树就不是最优工具。对于诸如此类的问题，我们可以采用其他更复杂的分类方法。

感知器和人工神经网络

决策树只是根据一系列实例推断（或学习）一般性分类规则的诸多技术之一。这些规则可以采取多种形式，也可以用多种不同的方式推断得到。例如，一种方式是将属性转换为数值（如果属性原本不是数值），然后对它们进行算术计算。一种显而易见并且受欢迎的方法是，根据某个标准确定输入属性的权重，计算加权和，然后与阈值进行比较。比较的结果可以确定该实例属于哪个类别。运用这种方法时，我们可以给不同的输入赋予不同的权重，就像我们在面对某个困难决策时会考虑不同变量的影响一样。我们在小心谨慎地选择权重后，算出加权和的值，就可以根据其类别标签的值来区分所有实例了。组合输入值时，也可以使用其他更复杂的形式。

计算属性值加权和的想法特别吸引人，一部分原因在于它的简单性，另一部分原因在于大脑中神经元的工作方式给了我们一种模糊的启发。沃伦·麦卡洛克和沃尔特·皮茨最早提出，大脑中的神经元对输入执行加权和计算，并将阈值函数的输出组合起来，去执行与思考有关的复杂任务。弗兰克·罗森布拉特于1958年提出感知器的概念，感知器是大脑神经元的简化模型。真正的神经元确实会对它们的输入进行计算，简单地说，这些计算就相当于求这些输入值的加权和。罗森布拉特感知器模型本来是作为硬件被设计出来的，它类似于人工神经元，因为它计算输入的加权和，并将计算结果与给定的阈值进行比较。如果这个值高于阈值，感知器就会被“触发”，它的输出为1；否则，输出为0。图5－3展示了一个简单的感知器执行的计算，有4个输入（x_1，x_2，x_3，x_4）和4个权重（w_1，w_2，w_3，w_4）。额外的权重w_0与一个值为1的常量输入有关，其实际作用是定义阈值。

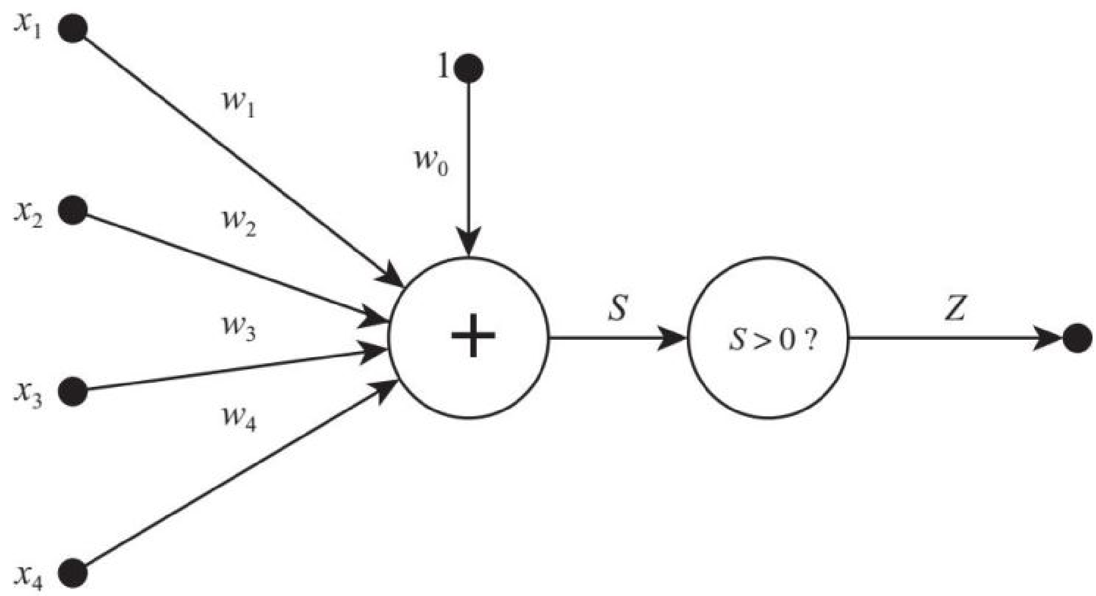


图5－3 感知器示意图

感知器激起了人们的极大兴趣，原因主要有两个。第一个原因是，在某些方面，它们与实际神经元的一般行为相似。真正的神经元也只在它们接收到的刺激总量很高时才会被触发，而在刺激总量较低时则会保持“沉默”。真正的神经元执行的计算比简单的求和复杂得多，在大多数情况下，计算结果也不太可能与输入的线性和一致。但是，如果我们想建立关于单个生物神经元行为的极简版数学模型，那么使用一个类似于感知器的模型是合情合理的。

第二个原因是，有一种学习算法可以用来训练感知器，还可以用来推导权重值和阈值。这个算法相当于一个数学法则，它调整感知器的输入权重，让感知器执行预期任务。这条法则很简单：每输入一个实例，如果预期输出Y与感知器的实际输出Z不一致，该法则就会修改每个权重。权重的变化幅度很小，受参数α控制，以保证加权和（在应用阈值之前）根据方程

$$w_i(t+1) \leftarrow w_i(t)+\alpha(Y-Z)x_i$$

向预期值靠拢。当输入x_i是正值时，这个方程可以表述为：如果预期输出Y大于观察到的输出Z，那么t+1时刻的权重w_i增大，反之则减小。如果x_i是负值，变化的方向则正好相反。该方程定义的变化表明，对于这个特定的输入模式，如果预期输出大于观察到的输出，加权和就会略微增大；反之，加权和将略微减小。这个更新的法则适用于所有权重w_i。只要满足一些非常简单的条件，这条法则就肯定能发挥作用，最终得到一组权重，从而使感知器可以对所有输入进行正确分类。

感知器模型似乎是朝着自重构系统的出现迈出的重要一步，自重构系统可以自主学习并执行预期任务。当时，基于感知器的学习被视为在计算机和电子系统中，开发出适应性类脑行为的一条前景光明的途径。事实上，1958年《纽约时报》乐观地称感知器是“一台处于萌芽阶段的电子计算机，（美国海军）期望这台计算机能行走、说话、视物、写字、自我复制，而且能意识到它自己的存在”。

遗憾的是，事情并没有这么简单。挑战来自于感知器的更新法则必须满足的条件中的一个，而且事实上，该法则只适用于单一感知器。感知器计算的是输入的简单加权和，在数学上这被称为线性函数，所以它只能区分线性可分的东西。例如，如果仅用两个实值属性x1和x2就能定义实例，那么这些实例可以被形象地表示成二维实空间中的点，每个属性对应一个维度。在这种简单的情况下，只在正实例和负实例可以被一条直线分开时，感知器才能将它们区分开。马文·明斯基和西蒙·派珀特在1969年出版的《感知器》一书中十分明确地指出，感知器无法学习分离更复杂的类别。基于感知器的方法在当时掀起的那股热潮被泼了一盆冷水，这促使研究人员为智能机器的开发寻找新的创意。

图5－4展示的是具有两个实值属性的假设性问题的两个例子。当只有两个属性时，问题可以被方便地绘制成二维图形，因为每个实例

分别对应二维平面上的一个点。图5－4（a）展示的问题是感知器可以学习的，因为利用一条直线就可以将正实例（实心圆）与负实例（空心圆）分开。而图5－4（b）展示的问题是感知器无法学习的，因为利用直线无法将这两类实例分开。表5－1所示问题的正实例（图5－1在表示这些实例时仅使用了三个维度中的两个）与负实例也可以被一条直线分开。因此，感知器可以学习这个问题。

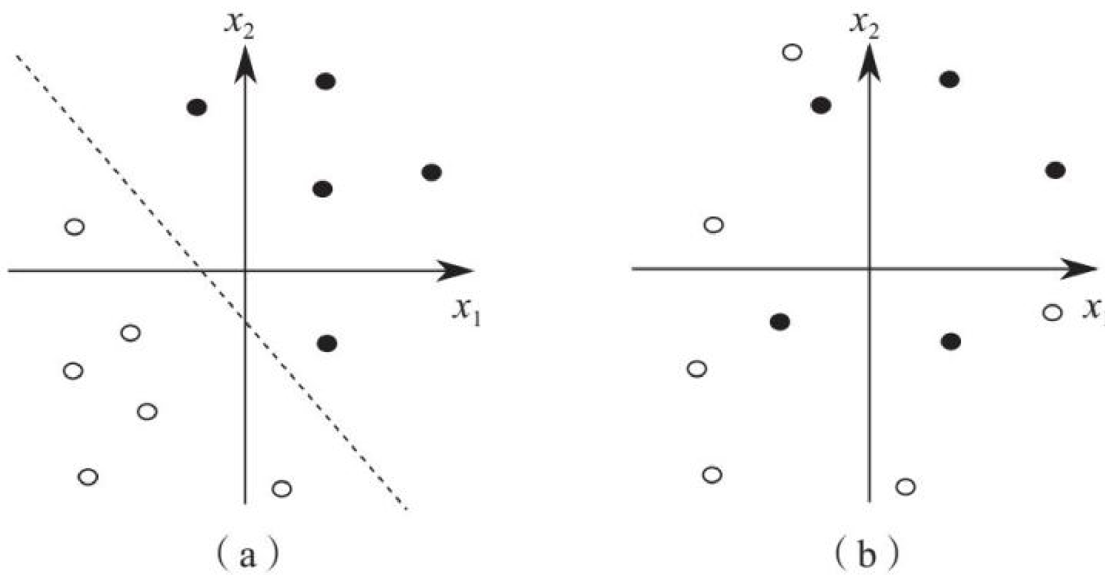


图5－4 两个不同问题的界线。感知器只能学习图（a）所示的问题，因为在这幅图中感知器执行的输入是线性函数

单一感知器无法学习更复杂的问题，这个消极结果本身并不会对基于感知器的方法造成严重打击，因为单一感知器毕竟只是单一神经元的模型，没人指望单一神经元可以执行智能任务。事实上，多项研究结果表明，多层人工神经网络可以十分近似地表现出任何预期行为，因此在正确地分配权重之后，就可以用来执行困难的任务。人工神经网络（通常简称为神经网络）是一种基于互连的简单神经元模型的计算模型。研究人员建立这些模型的初衷并不是如实地再现真实神经元的行为，而是把它们视为一种非常有用的抽象概念。

人们曾经希望利用某个学习法则去训练多层感知器网络。这种学习规则必须能解决所谓的信用分配问题，包括如何分配权重值，从而使整个网络可以执行需要完成的任务。

事实证明，信用分配问题很难解决。我们现在知道这是一个NP难度问题。多项研究结果表明，即使训练简单的感知器网络也是一个难解问题，这意味着没有任何已知的有效算法能确保在所有情况下都有效，并能为多层神经网络分配正确的权重。

但是，当能够训练软阈值多层感知器的算法被再次发现并得到普及（主要得益于戴维·鲁梅哈特、詹姆斯·麦克莱兰和PDP研究小组的研究成果）时，人们想找到一种有效的训练算法的愿望部分成真。这一突破的取得可能是因为，人们意识到如果将具有硬阈值的标准感知器替换为软阈值感知器，就有可能设计出一种训练算法，它至少在原则上可以推导出任意感知器网络所需的权重。不仅如此，这项成果甚至被扩展为有反馈的感知器网络——感知器的输出被直接或间接地反馈给它们自己。

这种算法为一种被称为连接机制的方法开辟了道路。尽管这种方法的发展速度比预期速度慢，但至今仍有很多人在使用它。连接主义者认为，由非常简单的处理器构建而成的网络，只要它的格局与权重是正确的，就可以用来执行非常复杂的任务，最终有可能表现出智能行为。

如果使用的是软阈值感知器网络（图5－5），输出神经元计算的就是一个平滑的良态函数。用数学术语来说，这意味着输入或连接权重的微小变化会导致输出也发生微小的变化。图5－5展示的例子中有5个输入单元、三个隐藏单元（在输入和输出之间）和一个输出单元。每个连接都有一个权重（未显示），调整权重就可以使多层感知器执行预期任务。

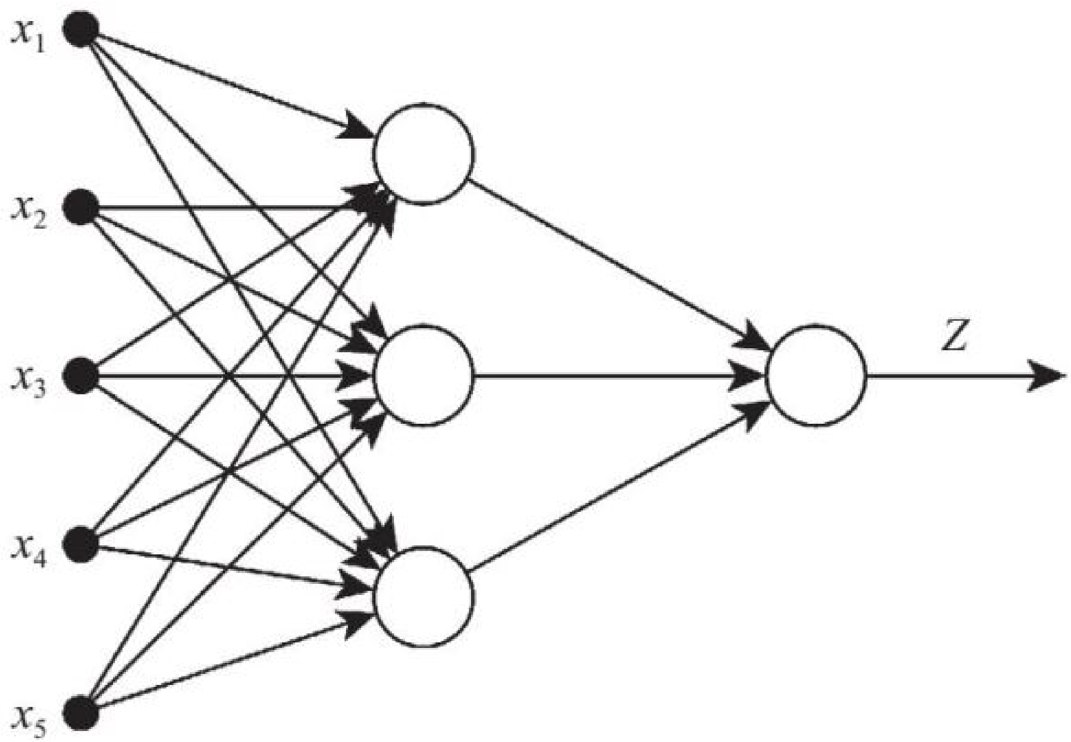


图5－5 包含输入层、中间层和输出层的多层感知器示意图

在数学上，对于任何可能的输入模式，都可以计算出每个互连权重的微小变化会使输出神经元的值发生怎样的变化。因此，我们很容易确定网络中任何权重的微小变化会如何影响输出。正因为如此，我们可以计算出应该如何改变权重，才能使输出神经元的值向预期方向靠拢。

更准确地说，我们计算的是输出误差对每个权重的导数。这些导数的集合被称为梯度，由于算法通过改变权重来使误差最小化，所以这种技巧被称为梯度下降法。反向传播算法是执行这种计算的一个简单数学公式，它先计算输出误差，然后反向传播至输入神经元，并在这个过程中得出误差对权重的完整梯度。之后，根据梯度调整权重，以减小输出误差。

从概念上讲，反向传播算法解决了信用分配问题，并证明了可用于训练感知器网络的有效算法是存在的——这似乎与认为训练多层感

知器是计算困难问题的理论相矛盾。这种明显的矛盾可以用下列事实来解释：尽管反向传播算法确实可以用来训练复杂的网络，但在许多情况下，它都找不到好的解决方案；也就是说，它找到的解决方案都不能保证误差很小，也不能保证网络学到的法则确实有用。

有些问题太过复杂，以至于不能直接通过反向传播算法来解决，比如，识别图像中的汽车或识别口语句子。这类问题需要先对输入值进行有效转换，派生出复杂的内部表征，然后将这些内部表征与以前的知识结合起来，从而得到最终答案。在这些情况下，反向传播算法的工作速度可能过慢，或者产生不合适的权重，因此无法解决信用分配问题。这是因为反向传播算法需要找出误差函数的最小值，但这很难做到。

为了调整将这些简单的类神经元单元（即感知器）连接起来的权重，人们发明了各种各样的技巧。除了反向传播算法以外，还有许多其他方法和模型，这使它成为机器学习领域中最复杂和研究最广泛的子领域之一。最近，一些统称为深度学习的新发展显著提升了神经网络的表现，并扩大了它的适用范围。深度学习法使用大量的训练数据、先进的多层体系结构和各种神经元模型，去推导包括复杂属性的中间表示在内的复杂结构。一些深度学习技术不仅受到标准的连接主义工具（比如反向传播算法）的启发，还从神经科学取得的进步，尤其是在真实的神经系统中观察到的模式中汲取灵感。

深度学习技术已被应用于计算机视觉、自动语音识别、游戏和自然语言处理等方面。事实证明，经过深度学习的训练后，神经网络产生的最新成果代表了该领域取得的重大进步。不过，包括深度学习在内的现有机器学习技术，最终都不太可能自主地推导出包含数十亿人工神经元，复杂程度堪比人类大脑的神经网络的完整连接。这些技术肯定会被用来执行汽车驾驶或人脸识别等复杂任务，但截至目前，这些任务仍需要人类的干预。我们必须取得更显著的进步，才有可能设计出通过机器学习最终表现出类人行为的智能系统。

贝叶斯公式

上文中描述的学习方法可能会给人一种特别安排的感觉，因为只有选择一个特定的框架，它们的参数才会适用于可用的数据。在实际执行任何学习之前，你必须选择将要使用的模型，而且该模型是根据现有数据推断出来的。例如，在构建分类器时，你可以选择决策树、神经网络或者数百种方法中的一种。只有在做出这个看起来相当随意的初步决定之后，你才能运用学习算法推导出分类器的结构和参数。这种方法背后的理论似乎不太可靠，因为从这个过程的一开始——学习范式的选择——就存在很大的随意性。

在实践中，人们会选择几种不同的方法并逐一尝试，然后选择效果最佳的那一种。这在一定程度上改进了学习过程，但仍有许多不尽如人意的地方，因为从选择待测试的分类器开始就表现出相当大的随意性，并且没有任何明确的法则可循。

有没有更好的方法呢？是否存在一个无须做出随意选择就能提供最佳答案，并得到适合所有未来实例的最佳分类法的数学公式呢？答案有点儿令人惊讶："是，也不是。"托马斯·贝叶斯最早发现了这个问题的答案。1763年，即他去世的两年后，他的一篇论文在英国皇家学会会议上被宣读，这个问题的答案就在其中。由理查德·普莱斯代贝叶斯提交的这篇论文题为《机会问题的解法》，它描述的是有限制条件的贝叶斯定理。现代版本的贝叶斯定理是数学家皮埃尔-西蒙·拉普拉斯进一步发展的结果，我们现在使用的贝叶斯公式最早发表于拉普拉斯的著作《分析概率论》中。

贝叶斯定理可以根据对某个事件有影响的其他事件的概率，计算出该事件的概率。具体来说，如果我们知道定义某个实例的属性值，

也知道该实例的标签取其他几个值的概率，就可以利用贝叶斯定理预测出该实例的标签取某个特定值的概率，而且这种预测具有数学确定性。已知两个相互影响的事件（A和B），贝叶斯定理可以简单地表述为：

$$P(A\mid B)=\frac{P(A)\times P(B\mid A)}{P(B)}。$$

利用这个公式，我们可以算出事件B发生后事件A的发生概率P(A ｜ B)。这个概率（即在事件B已经发生的条件下事件A的发生概率）在贝叶斯统计学中具有非常重要的意义。我将借助表5－1中的数据来说明它的作用。

我们把事件A定义为“适合打网球的日子”，把事件B定义为“风速小于15节[注]的没有风的日子”。如果我们假设这些数据代表了天气特征的分布，并且可以如实地再现事件的概率，就可以得出P(A)=3/7，因为在有记录的7天中只有3天是适合打网球的日子。但是，如果我们只考虑没有风的日子，P(A ｜ B)的值就为3/5，这是在无风条件下适合打网球的日子的概率。

在将这个公式应用于归纳式学习问题之前，我们有必要考虑一个更简单的例子，看看如何正确地运用贝叶斯公式来推理概率。

假设在没有任何特殊理由的前提下，随机地从所有美国人中选择一个人接受艾滋病检测。他找到的那家诊所声誉良好，检测准确率为98%。这意味着它给出假阳性或假阴性的错误结果的概率只有2%。我们知道，美国的艾滋病发病率估计为0.4%，这意味着在随机抽取的1 000人样本中，大约有4人患有艾滋病。

现在，我们假设检测结果是阳性。那么，这个人患有艾滋病的概率是多少呢？考虑到这种检测的错误率很低，有人可能会认为这个人患有艾滋病的概率相当高。这个人患有艾滋病的确切概率可以用贝叶斯定理计算得到，其中A表示“患有艾滋病”，B表示“检测结果为艾滋病阳性”。那么，在检测结果为阳性的情况下，$P(A \mid B)$和$P(\overline{A} \mid B)$（分别表示这个人患有艾滋病的概率和未患艾滋病的概率）的值是多少呢？

前文说过，美国的艾滋病发病概率$P(A)$为0.004。因此，$P(\overline{A})=0.996$。由检测结果的误差概率可知，$P(B \mid A)=0.98$，$P(B \mid \overline{A})=0.02$。这两个值分别代表检测结果为真阳性（这个人患有艾滋病且检测结果为阳性）的概率和检测结果为假阳性（这个人未患艾滋病但检测结果为阳性）的概率。

通过计算可知，$P(A \mid B)=P(A)\times P(B \mid A)$和$P(\overline{A} \mid B)=P(\overline{A})\times P(B \mid \overline{A})$的值分别为0.0039和0.020。在这两种情况下，贝叶斯公式的分母中的归一化因子$P(B)$相同，它唯一的作用就是使这两个值的和为1，因为这个人要么患有艾滋病，要么未患艾滋病。在对0.0039和0.020这两个值进行归一化处理，从而使它们的和为1之后，我们将会得到一个反直觉的结果：即使在艾滋病检测结果呈阳性的情况下，这个人患有艾滋病的概率也只有16%，而未患艾滋病的概率为84%。

贝叶斯定理可以使我们将事件的先验概率与改变我们对这些事件发生概率的置信度的额外证据（在本例中指检测结果）结合起来，从而得到后验概率。例如，地震较为罕见，因此当你感到建筑物的地板在晃动时，你首先会想到其他原因。但是，如果你看到整座建筑物都在摇晃，这个新增的证据就会盖过地震的较低发生概率，让你认为可能正在发生地震。

我们将贝叶斯公式应用于表5－1所示的例子。假设事件$\overline{A}$代表“适合打网球的日子”，事件A代表“不适合打网球的日子”。我们先考虑这个问题的简化版本，只用表5－1第一列中的数据定义事件B。事件B意味着温暖的日子（在本例中，我们将其定义为温度高于10℃的日子），事件B意味着寒冷的日子（温度低于10℃）。现在，我们可以计算出应用贝叶斯定理需要的所有值。

接下来，我们可利用该表估算出以下概率：P(A)=3/7，因为7天中有3天是适合打网球的日子；P(B ｜ A)=1.0，因为适合打网球的日子都很温暖；P(B)=5/7，因为7天中有5天是温暖的日子。有了这些概率，我们就可以运用贝叶斯定理算出天气温暖且适合打网球的日子的概率P(A ｜ B)。通过下式

$$\frac{(3/7)\times 1.0}{5/7}$$

我们算出P(A ｜ B)为0.6。观察图5－1，温暖的日子一共有5天，其中有3天适合打网球，由此证明我们算出的概率是正确的。

请注意，这个计算不需要专门假设什么样的日子适合打网球。我们只是基于数据计算概率，然后应用贝叶斯定理得出确切的结果。但是，贝叶斯定理在现实生活中的应用面临着重重困难，因此几乎无法直接使用。这些困难与贝叶斯公式右边的取值所需的概率计算有关。

例如，P(B)的值，即温暖日子的概率就很难计算。我们根据表5－1估算出这个值为5/7，但我们无法保证对P(B)的真实值的估算是准确的。我们可以通过收集更多数据或查阅气象记录来改善这个估算值，从而更准确地估计P(B)的真实值。

但是，如果要考虑表5－1中所有列的数据，就会面临更大的困难。假设我们要根据同时发生的其他事件来计算事件A的概率。例如，我们可能想计算P(A ｜ B∪C∪D)，其中B表示温暖的日子，C表示干燥的日子，D表示有风的日子。记住，符号“∪”的意思是“并”，表示

这些属性值同时出现。就像之前一样，要应用贝叶斯定理，不仅需要计算P(A)的值，还需要计算P(B∪CD ｜ A)的值。后一个概率很难计算，因此增加了贝叶斯定理的应用难度。P(B∪C∪D ｜ A)是具备这些特征（温暖、干燥、有风）且适合打网球的日子的概率。精确计算这个概率需要查阅大量记录，这在许多情况下是做不到的。即使有气象记录也无济于事，除非这些记录还包括这些日子是否适合打网球的信息。如果我们要处理的不是三个简单的属性，而是几百个属性，那么我们永远无法获得这类信息，因为我们之前可能从未考虑把这些属性组合到一起。

因此，我们通常不能直接应用贝叶斯定理来计算某个事件的“真实”概率。但是，我们可以用多种方法完成近似计算，许多实用的分类器和学习方法由此产生。有一种叫作朴素贝叶斯分类器的简单方法认为，对于给定的分类，我们假设属性的值相互独立，然后使用“朴素”的近似值取代难以计算的概率P(B∪C∪D ｜ A)。这意味着P(B∪C∪D ｜ A)可以由P(B ｜ A)×P(C ｜ A)×P(D ｜ A)来代替，后面这个值很容易计算，因为它的每一个因数都能轻易地根据实例表格或其他记录计算出来。在这个例子中，它的假设相当于，在适合打网球的日子里，天气温暖、干燥且有风的概率，可以通过天气温暖的概率乘以天气干燥的概率再乘以有风的概率计算出来。坦率地说，这些概率相乘通常不能得到正确的结果。为简单起见，我们假设适合打网球的日子中有一半是温暖的天气，有一半是干燥的天气，有一半是有风的天气。那么，它意味着在适合打网球的日子中天气温暖、干燥且有风的概率是1/8，因为P(B∪C∪D ｜ A)近似等于P(B ｜ A)×P(C ｜ A)×P(D ｜ A)。这个假设可能是不正确的，因为温度、干燥度和有风天气之间具有相关性，所以温暖、干燥和有风天气的概率不能简单地用它们三个的概率相乘来计算近似值。

尽管朴素贝叶斯分类器使用的简化假设在许多情况下都是不正确的，但这种分类器很强大，可以解决很多问题；而且在许多情况下，与那些更复杂的方法相比，朴素贝叶斯分类器的效果好得惊人。

大脑、统计学和学习

在机器学习中运用统计学去推断如何正确应对从未见过的情况，似乎会让人产生一种人为设计的感觉，与生物系统的做法存在很大的区别。但令人惊讶的是，事实并非如此。

假设有一张与表5－1类似的表格，但天气属性换成了其他属性，标签值也换成了其他行动方案。这样的表格可以代表更复杂的问题（可能是生死攸关的问题），如表5－2所示。在这种情况下，选择何种行动关乎个体的生死存亡。

表5–2 假设性的早期人类决策问题

狮子	野兔	豹子	大象	树	行动
可见	附近	可见	未见	附近	逃跑
可见	遥远	未见	未见	遥远	逃跑
未见	附近	未见	可见	遥远	追捕
未见	附近	未见	未见	附近	追捕

现在人们认为，大脑实际上是一部非常复杂的统计机器。大脑中存有外部世界的相关部分的内部模型，并利用该模型估算贝叶斯定理涉及的概率，然后采取适当的行动。

大脑使用的概率模型比前文中讨论的模型要复杂得多，而且大脑评估的对象并不局限于静态属性定义的实例。将时间纳入考虑范畴，并可以为一系列行动的结果创建内部模型的机器学习技术，已经在许多机器学习应用程序、机器人技术、游戏和优化程序中得到了广泛应

用。强化学习，尤其是一种被称为时序差分学习的强化学习形式，就是这样一种技术。强化学习的思想基础是，每个特定的行动都有一个期望奖励。这种奖励可以根据多个行动－奖励对估算出来，即使这些奖励相对于这些行动有明显的延迟。每种生物的大脑中很可能都存在这种可获得类似结果的估算过程，从而创建出评估所有行动结果并据此选择最佳行动的能力。

认为大脑本质上是一部非常复杂的统计机器的概念，可能是肯尼斯·克雷克提出的。1967年，克雷克在他的《解释的本质》一书中提出，生物的大脑中装载了现实世界的“小尺度模型”，可以捕捉外部事件之间的统计关系。了解大脑如何处理感官信息并在其内部构建外部世界的模型，一直是神经科学追求的核心目标之一。但是，经过数十年的研究，这个目标仍未达成。

人脑拥有计算复杂的统计信息的能力，是所有已知大脑中功能最强大的，它可能是由一个持续了数十亿年的过程精心设计出来的一个极其复杂的产物。从某种意义上说，查尔斯·达尔文发现的这个过程是有史以来最复杂和运行时间最长的算法，它的名字叫作进化。

---

1. 1节=1.852千米/时。——编者注

第6章

细胞、身体和大脑

查尔斯·达尔文在撰写《物种起源》之前，他的研究初衷并不是掀起一场空前绝后的科学革命，也不是改变人们对世界的看法，使之完成有史以来最深刻的范式转移。

在人类历史上，没有哪一种思想能像达尔文的理论那样，对我们今天的思考方式产生如此大的影响。虽然许多其他重要的范式转移也深刻地改变了人们对世界的看法，包括哥白尼对地心说的颠覆、牛顿运动定律、爱因斯坦的相对论、原子理论、量子力学、内燃机和集成电路等；但是，这些突破都未曾像达尔文进化论那样深刻地改变了我

们对世界和我们自己的看法，而达尔文的初衷只是解释为什么地球上有这么多不同的物种。

终极算法

达尔文早年对大自然的兴趣点燃了他对自然科学的热情。在他乘坐贝格尔号航行的5年里，这份热情达到了顶点。航海日志的出版，使达尔文成为一名受欢迎的著名作家。因为对野生动物的地理分布和在航程中收集到的化石种类感到迷惑不解，达尔文提出了自然选择理论。包括阿尔弗雷德·华莱士在内的其他科学家也得出了类似的结论，但达尔文是第一个发表这些结论的人。事实上，当达尔文了解到华莱士的想法后，他不得不匆忙完成并出版了自己的著作，而在此之前，他的研究一直保持着一种闲适的步调。

达尔文在书中描述了一个简洁有力的观点。他认为，之所以会出现不同的物种，是因为通过选择优势特质，选择压力对个体的繁殖成效产生了不同的影响，即使只是在很小的程度上：

> 如果自然界中真的有变异，还有一股强大的力量随时会发挥选择作用，那么只要这些变异对生物是有利的，即便它们与生命的关系极度复杂，我们又凭什么怀疑这些变异不会被保存、累积并传承呢？人既然能耐心选择对自己有用的变异，大自然为什么不能选择在变化的生命条件下有利于生物的变异呢？对于这种在漫长的岁月中发生作用并严格检视每一种生物的整体组成、结构和习性——取其精华，去其糟粕——的力量，我们能加以限制吗？对于这种缓慢并美妙地使每一种生物适应最复杂生命关系的力量，我看不到有什么能限制它。即使我们无法看得更远，自然选择理论在我眼里也是可信的。

达尔文的核心观点是，由于每一代个体都可以生育多个后代，所以物种的繁殖倾向于遵循指数定律。达尔文知道，如果自然选择不限制种群的增长，那么即使单一物种以指数速度增长，它们也会遍布整个地球，并耗尽所有可用资源。物种之间的竞争可以确保适应环境能力强的物种在后代的存活数量上具有竞争优势。随着时间的推移，具有不同特征的种群在增长速度上的差异（可能很小），推动了优势性状的选择。孔雀的羽毛、海豚的尾巴、长颈鹿的脖子和人类的大脑都适用这个原则。

该原则也可以用来解释我们在人类身上看到的所有特征，这对认为人类物种在地球上（甚至是宇宙中）起着至关重要作用的主流观点来说，具有不小的杀伤力。在《物种起源》中，达尔文甚至没有明确提出这一观点。他只是顺便提到“人类的起源和历史终将得以阐明”，也就是说，他认识到人类与其他物种没有什么不同，因为人类也是在自然选择的压力下进化产生的。

达尔文非常清楚地阐述了自然选择进化的概念，并且在150年的科学发展历程中保持不变，这本身就非常了不起。我们现在知道，只有当遗传的离散单元能以某种方式在后代的身体中进行自我复制时，达尔文提出的自然选择进化才有可能发生。这些单元——我们现在称之为基因——代代相传，几乎不会发生任何变化。

更值得一提的是，达尔文得出的结论至今仍然有效，但却没有提到特定特征由亲本遗传给后代的基本机制。因为达尔文不知道他同时代的格雷戈尔·孟德尔的研究成果，所以他几乎没有考虑过基因遗传背后的物理机制，仅有的一点儿想法也是错误的，尽管基因是进化理论必不可少的遗传离散单位。如果物理特征是通过连续的模拟编码从亲本传递给后代的，进化过程就不可能迈出它的第一步。

在大多数物种中，一代又一代的个体通过交配繁衍生息。因此，即使身高等特定特征具有选择性优势，但如果个体的基因组成不是以独立的方式编码，这种优势也会随着连续几代的繁衍而减弱。要理解

其中的道理，可以想象一组玻璃杯，每个都装满了染色的水，但染料的量不等。现在假设选择机制开始发挥作用，并选择了水颜色较深的玻璃杯。为了获得下一代，水颜色较深的玻璃杯被选中。经过混合，下一代水的颜色肯定比上一代深，而且从总体上看，所有玻璃杯里的水颜色都比原来深。但是，由于下一代的水是通过混合的方式得到的，所以它们的颜色分布比上一代均匀得多。随着世代的推移，所有杯子里的水都会变成相同的颜色，并且不再加深。

如果用独立的编码记录每个玻璃杯中的染料量，情况则会迥然不同，传递给后代的可能就是这个独立的数字，而不是染色的水本身。这个办法可以避免种群向最优秀个体的平均水平收敛的问题，因为即使经过许多代的遗传，群体中仍然存在着不同的染料浓度，甚至最高的浓度也被保留下来。不仅如此，这还为一些罕见但并非不可能发生的事件（比如突变），以及无止境的开放式进化创造了条件。

孟德尔的研究揭示了基因遗传的真正机制。孟德尔是遗传学之父，为了接受教育，他成为奥古斯丁修道院的一名修士。为了研究植物的变异，他在1856—1863年种植了大约29 000株豌豆。他选择对7种植物特征进行研究，其中包括种子形态、荚果形态、种子颜色、荚果颜色和花的颜色等。我们现在知道，孟德尔模型体现的这些性状之所以能保持独立性，是因为对这些性状进行编码的基因位于不同的染色体上。

孟德尔依据他的大量实验结果，找到了现在被称为孟德尔遗传定律的机制。1866年，他的研究成果发表在一份不起眼也没有名气的期刊上，孟德尔在文章中指出，遗传性状以一种数字开/关的形式传递给后代；而且，他研究的生物（豌豆植株）的每个基因都有两个拷贝，其中一个被传递给下一代。事实上，孟德尔的实验结果适用于所有二倍体生物（每条染色体均有两个拷贝的生物）。大多数动植物都是二倍体，但也有一些是多倍体，例如，鲑鱼的每条染色体都有三个拷贝。

孟德尔进一步推断出，不同的性状是彼此独立地传递给下一代的。他研究的性状的确如此，但总的来说，这只适用于由独立基因编码的性状。所以，我们有必要更具体地讨论他的研究结果。

假设一种植物的某个性状是由等位基因编码的。等位基因有两个值，它只能取其中一个：A（假设它是显性的）或a（假设它是隐性的）。如果等位基因的一个值遮蔽了另一个值，我们就说前者是显性的。在本例中，我们假设每个基因只有两个等位基因，每条染色体有两个拷贝，因此每个基因座有4个可能的值：AA，Aa，aA和aa。由于等位基因A是显性的，所以具有AA、Aa、aA变体的植株会表现出该性状，而具有aa变体的植株则不会表现出该性状。当植物繁殖时，后代从两个亲本那里各继承一个基因拷贝。有了这些信息，我们就可以继续讨论孟德尔的分析——现代遗传学的基础。

我们先用一些具有AA变体的植株和一些具有aa变体的植株做实验。将这两种植株杂交，就会得到Aa植株。Aa植株与它们的亲本AA一样，也会表现出这种性状，因为等位基因A统治着等位基因a。如果将Aa植株与其他AA植株杂交，第二代的情况就会变得更有趣。由于下一代将从两个亲本处各继承一个基因，所以我们会得到AA植株（约占25%）、Aa和aA植株（约占50%），以及aa植株（约占25%）。由于AA和Aa植株均表现出这种性状，而aa植株不表现出这种性状，所以我们预计表现出这种性状的植株和不表现出这种性状的植株之比为3∶1。这与孟德尔得到的实验结果恰好一致。

孟德尔通过他精心设计的实验步骤，令人信服地验证了他的假说：性状是通过结合亲本的基因以独立的方式传递给下一代的。事实上，他得出的结论如此清晰、精确和令人信服，以至于多年后有人围绕他的实验结果的极高精确度是否可信的问题展开了争论。统计学家和遗传学家罗纳德·费舍尔指出，孟德尔从他的实验中得到的实验比率与3∶1这个理想比率极为接近，简直令人难以置信。费舍尔分析了在第二代中产生3∶1的比率的那些实验结果，他发现如果重新进行实

验，那么实验结果更接近这个理论比率的概率将不到千分之一。因此，要么是孟德尔极其幸运，要么是他的实验结果有问题。现在出版的很多书籍或发表的许多文章仍在讨论这个问题，争论远未结束。尽管少有人指责孟德尔篡改数据，但对许多人来说，这些实验结果仍然很神秘。其中有一种观点认为，当某些实验结果与孟德尔的假设不完全一致时，他可能舍弃了这些实验结果，或者重新进行实验。

孟德尔的研究在他那个时代基本上被忽视了。直到他去世多年后，人们才发现必须用离散模型来解释遗传机制，并广泛接受了他的研究成果。1953年，詹姆斯·沃森和弗朗西斯·克里克确定DNA是负责遗传的分子，并发现了它的螺旋结构，这才从物理学和分子学角度解释了孟德尔的实验结果。

由于当时的已知机制都无法为进化所需的遗传类型提供支撑，所以达尔文的理论遇到了一些小麻烦。尽管如此，进化论在随后的所有讨论中都经受住了考验。用今天的术语来说，达尔文认为生命的多样性是一个在地球上运行了大约40亿年的盲算法作用的结果。对现在熟悉算法概念的读者来说，把进化看作一种算法，丝毫不会让人觉得意外。丹尼尔·丹尼特清楚而完美地表达了这一观点：

> 达尔文的抽象想法的理论力量来自他确定的几个特征，而且他对这些特征的理解比他的许多支持者更深刻，但他没有用术语对它们进行明确的描述。今天，我们可以用一个术语来描述这些特征，那就是达尔文发现了算法的力量。

地球上所有形式的生命都是被一种算法创造出来的，这个过程不加思考地应用固定不变的法则，逐步派生出今天这个极其复杂的世界，这种想法让人们大吃一惊。毕竟，关于不同的物种如何出现在地球上的古老问题，有一个简单且不言自明的答案。自然选择足以解释生命树中的物种多样性。公平地说，许多问题仍然悬而未决。无论是

过去还是现在，进化如何开始的相关问题始终最令人头疼。毕竟，只在繁殖方案已经确定，以及有生物繁殖和发生变异的前提下，我们才可以用自然选择理论解释繁殖速度各不相同的问题。达尔文没有尝试回答这个问题。即使到了今天，关于世界上第一类可以某种方式复制自己的系统，我们也只有一点儿非常模糊的想法。

很多人反对进化论，还有很多人对它持怀疑态度。神创论认为，生命是由神圣实体创造出来的，也许就在几千年前，而且它们在诞生之初就是现在的形态。在与进化论针锋相对的众多观点中，最强有力的一个是，无论操控这个过程的盲算法有多么精妙，一系列随机的变化都不可能产生我们今天看到的这些复杂的设计。该观点还认为，在几十亿年的时间长河中，某个睿智的设计师于某一时刻参与其中。如果我们看到一块手表（一个复杂的装置），我们自然会推断出有钟表匠存在；然而，当我们看到一只眼睛（比手表复杂得多的东西）时，却为何会认为它是由一些小变化以几乎完全随机的方式日积月累地进化而来？许多作者对此给出了完美的回答，但在这里我只能引用其中最有说服力的一位——理查德·道金斯的观点：

> 自然选择就是那位盲眼钟表匠。说它盲，是因为它从不展望未来，不计后果，也不会锚定任何目标。但自然选择就像一位杰出的钟表匠，它创造出的生动结果以其外观设计给我们留下了不可磨灭的印象，并让我们产生一种精心设计、细心谋划的错觉。

不过，尽管我们中的大多数人都承认进化是一个盲选过程，它创造出所有大大小小的生物，但在某种程度上，我们仍会认为这个过程的最终产物必然是智人。事实很可能不是这样，进化的目的并不是为了创造出任何特别的东西。如果这一过程从头再来一遍，就极有可能不会出现像人类这样的物种。在复杂生物的进化过程中，有太多的意外、太多极不可能发生的事件，以及太多的随机性。有无数事实说明历史事件完全有可能产生不同的结局：第一种可自我繁殖的实体的出

现，真核细胞的进化，寒武纪大爆发，恐龙灭绝，甚至是智人在经历诸多严峻的挑战和进化的瓶颈后幸运地存活下来（他们的表亲就没有这么幸运了），等等。如果这些事件没有发生，或者以一种迥异的方式结束，地球的生命历史将完全不同，现在居住在地球上的可能就是迥然不同的物种。用著名进化生物学家古尔德的话说：

> 我们是一个不可预测、不受复杂性驱动的过程在不经意间创造的完美产物，而不是渴望产生一种能了解其必要构造的生物的进化原则的预期结果。

同样的道理，这个历时数十亿年的过程最终也不太可能产生智人这个物种。但是，在地球的历史上，确实有一个物种第一次掌握了控制进化的工具。几百年来，我们一直对作物和狗、马等家畜实施人力驱动的选择，并通过这种方式控制了进化。但今天，我们即将开发出可以直接控制地球上几乎所有生命形式的繁殖成效的技术，我们甚至有可能创造出新的生命形式。

有必要提醒大家的是，自然选择并不会评估和奖励个体的适合度。的确，后代的数量与特定个体的适合度存在着统计相关性，但考虑到基因遗传的机制和孟德尔图书馆的规模，即使适合度强的个体或许也只能在整个宇宙历史上存在一次。自然选择不会选择杰出个体的未来拷贝，为未来的轮回转世做准备。一般来说，至少在有性繁殖的生物中，即使是最成功的个体，其后代也会大不同。相反，基因这种难以捉摸的实体的特定变体将被遗传给后代。虽然我们在直觉上可能会认为自然选择是在选择个体，但更合理的做法是把基因看作选择的基本单位。

基因是原始的实体，在大约40亿年前，它们找到了自我复制的方法。因此，它们变成了复制因子。成功的基因变得更常见，如果选择压力不抑制它们的扩张，它们的数量就会以指数速度增长。结果是，

一个种群中最常见的基因平均来看就是那些通过大量基因型并在很多情况下对其宿主产生最有利影响的基因。换句话说，我们应该预见到自私基因的蓬勃发展，即它们会促进自身的生存，而不一定会促进生物、群体甚至是物种的生存。这意味着进化适应是基因对其宿主施加的影响，目的是在后代中最大化地表达自己。如果一种适应能直接促进宿主的生存，或者它能通过促进其他目标最终帮助宿主成功繁殖，这种适应就会被选中。

随着时间的推移，为了生存下去，这些基因开始为越来越复杂的结构编码：先是简单的细胞，然后是复杂的真核细胞，再然后是身体，最后是大脑。事实上，正如我们今天知道的那样，生物的存在主要是为了确保这些单位（基因）的有效复制。最极端的观点认为，生物存在的唯一目的就是维护DNA的利益，确保DNA分子被传递给下一代。基因——编码特定蛋白质或控制特定过程的DNA序列——的寿命是以百万年为单位衡量的。一个基因可以延续数千代，而且几乎没有变化。从这个意义上说，生物个体不过是一代代传递下来的DNA信息的临时载体。

我们可能会认为，随着时间的流逝，现在情况已经发生了改变，进化的中心是生物，而不是我们已知的以DNA碱基的特定组合形式编码的遗传单位。或者说，这些原始的复制因子已经消失，取而代之的是复杂的生物，但这种想法太天真了。或许在人类从地球上消失的数百万年后，最有效的复制因子仍存在于这个星球上，甚至是已知宇宙中。我们无法知道这些复制因子在数百万年后的样子。它们可能与今天的基因非常相似，仍在为细胞的特定特征编码，这些细胞形成的生物与今天存在的生物也非常相似。但是，我不这样认为。我非常认同道金斯的一段话，它清楚地表明我们不过是基因生存的载体，我们应该意识到基因确实是足智多谋的实体：

> 复制因子为了确保它们自身在地球上的存续，不断改进它们采取的各种技术和手段，这种改进有没有止境呢？它们有大把的时间

来实现这些改进。几千年的时间会产生何种怪诞的自我保护引擎呢？40亿年后，远古的复制因子又将面临何种命运呢？它们没有灭绝，因为它们是昔日的生存艺术大师。但是，不要以为它们还漂浮在海洋之中，它们很久以前就放弃了这种自由自在的生活方式。现在，它们过起了群居生活，安全地待在巨大而笨重的“机器人”体内，与外界隔绝，通过迂回曲折的间接路线与外部世界交流，并用遥控的方式操纵它。它们存在于你我之中，它们创造了我们的身体和心灵，它们的存续是我们生存的终极理由。这些复制因子已经走过了一段漫漫长路。现在它们被称为基因，我们是它们的生存机器。

考虑到以固定速度繁殖的实体的指数增长特征，我们可以肯定的是，复制因子在几百万年后仍将以某种形式存在。很多复制因子可能不再将DNA用作编码信息的底物，但肯定还有一些复制因子会继续使用这种可靠的方法。

复制因子的本质不是存储信息的媒介，而是被存储的信息。只要复制实体的机制可用，任何东西都可以充当底物。一开始，DNA可能不是底物的首选，因为它需要用到当时还不存在的复杂繁殖机制。事实上，还有一种分子被广泛地用来存储和传输细胞中的遗传信息，即RNA（核糖核酸）。RNA可能在DNA之前被用作存储生物信息的底物，在细胞内部起着多种生物学作用，与DNA不同，RNA通常是单链结构。

根据RNA世界假说，在以DNA为基础的生命形式出现之前，RNA分子是地球上所有现存生命的早期形式。经过广泛的讨论，这个假说受到了许多研究人员的青睐，因为有证据表明，诸多年代久远且非常稳定的细胞结构都基于RNA。不过，即使是RNA世界假说的支持者也认为，在RNA之前可能还存在其他不太发达的底物。但直到现在，这个问题仍是一个未解之谜。

未来，也许会有比DNA更善于存储和处理复制因子生存繁衍所需信息的其他底物。我们都熟悉计算机病毒，知道它们在合适的条件下会

以指数方式进行自我复制。不过，这些计算机病毒是非常原始的复制因子，简单且脆弱。未来，计算机存储器和中央处理器中也许会出现能自我复制的更复杂实体，它们可能是未来的复制因子。今天的生物与40亿年前在地球上的温暖池塘中繁衍生息的复制因子相似，未来的复制因子也将与今天的基于DNA的生物相似。

细胞和基因组

在了解了自然选择进化的算法，并且知道了基因这种带有些许神秘色彩的遗传单位之后，我们就可以进一步理解昔日的基本复制因子（基因）是如何进化出它们现在赖以生存的复杂结构（多细胞生物）的。

即使在达尔文的研究成果被发表和接受之后，生物学仍然是一个比较平静的非热门研究领域。自然选择进化的原理和遗传学的基本原则，为生物学成为一门以第一原理为基础的科学打开了一扇门。但在很长一段时间里，由于人们不了解支撑生命的物理机制和遗传学，所以真正意义上的生物学原理研究一直举步维艰。事实上，在20世纪的大部分时间里，由于人们对生命的生化基础不甚了解，对遗传特征代代相传的方式也知之甚少，所以进化和物种形成的基本机制一直是个未解之谜。直到1953年，还没有人确切地知道遗传特征是如何从亲本传递给后代的。

在达尔文致力于进化论研究的同时，许多生物学家把目光投向了细胞在生物学领域的核心地位。1838年，西奥多·施旺和马蒂亚斯·施莱登首次提出生物是由细胞构成的，以及单个细胞具有活性的观点。到1850年前后，大多数生物学家都逐渐接受了细胞是生物的基本单位，以及细胞本身具有活性且能繁殖的观点。到1900年，人们发现了很多代谢途径，即分子相互作用的方式。以色谱法（基于凝胶的混合分离技术）和电泳（利用电磁场分离不同的带电粒子）为代表的先进技术促进了生物化学的发展，并使人们对在细胞中发现的一些基本的生化机制有了一定程度的理解。

奥斯瓦尔德·埃弗里、科林·麦克劳德和麦克林·麦卡蒂的研究证明，DNA是导致某些致病菌产生可遗传变化的物质，他们指出DNA负责将遗传信息传递给后代。但是，直到1953年沃森和克里克证明DNA分子是遗传信息的载体，遗传学才开始发挥它现在起到的核心作用。在此之前，人们对DNA的作用几乎一无所知。生物物理学家麦克斯·德尔布吕克甚至把DNA称作“愚蠢的分子”，因为它总是单调地重复着4个碱基A、C、T和G（分别代表腺嘌呤、胞嘧啶、胸腺嘧啶和鸟嘌呤），而人们已经知道蛋白质具有多种多样的形式和功能。沃森和克里克利用罗莎琳德·富兰克林和莫里斯·威尔金斯收集的X射线数据，提出了DNA分子的双螺旋结构，从而解决了一个存在已久的问题：基因信息是如何从一个细胞传递给另一个细胞，以及如何从亲本传递给后代的？

沃森和克里克发现，DNA的双螺旋结构是遗传信息的储存库，遗传信息以数字格式存储在DNA分子中。而在此之前，他们并没有想到这将是一个令人兴奋不已的发现，它翻开了生物学历史上一个跨越几十年的最激动人心的篇章。突然间，孟德尔的思想和达尔文的进化论固有的独立遗传编码机制被发现了，其形式是一个具有简单重复结构的分子。DNA分子双螺旋结构的每个单链的每个片段都包含一个碱基或核苷酸，为A、C、T、G这4个可能值中的一个编码。双链之间的化学键的性质可以保证A总是与T配对，C总是与G配对，如图6－1所示。这种丰余为DNA分子创建两个拷贝提供了条件：双螺旋先分裂成两个单螺旋，然后以这两个单螺旋为模板，创建出两个新的双链DNA分子。

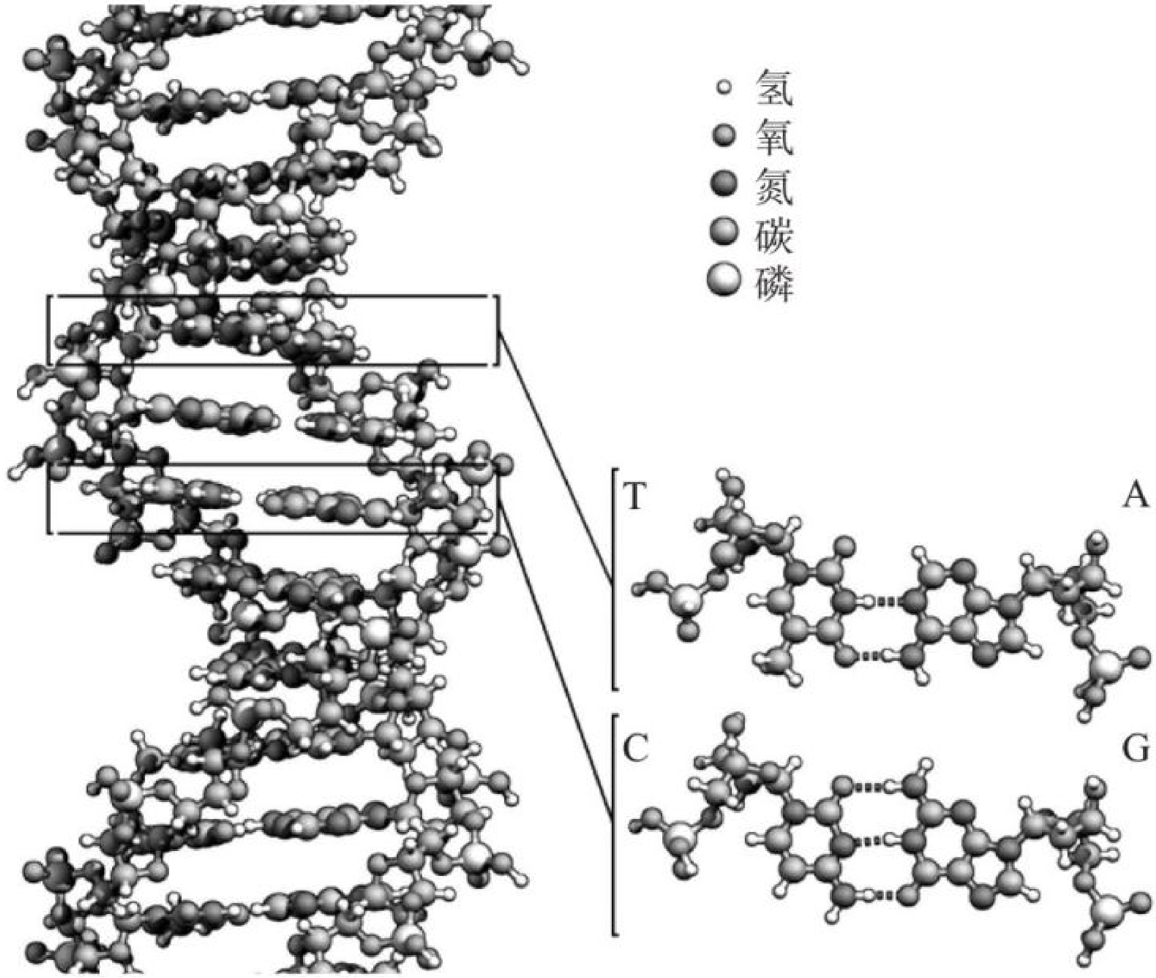


图6－1 DNA分子的双螺旋结构，其中A－T和C－G碱基对为细胞分裂或生物繁殖创造了DNA复制所需的丰余
图片来源：理查德·惠勒。

对于大自然如何将特征从一个细胞传递给另一个细胞，以及从亲本传递给后代，我们在理解方面取得了巨大进步，最终引发了另一场革命，并且持续至今。伴随着这一发现，人们终于明白了当今世界的复制因子是由DNA构成的，细胞的所有活动也都是围绕这种分子进行的。DNA是细胞分裂时运载遗传信息并将这些信息从亲本传递给后代的载体。因此，DNA肯定包含以其长碱基序列编码的作为遗传基本单位的结构。

我们现在知道，DNA以各种各样的方式对细胞中使用的信息进行编码，但对其中很多方式我们都一知半解。不过，DNA最重要的作用是编

码蛋白质中的氨基酸序列。蛋白质是氨基酸序列，一旦形成，就会被折叠成复杂的三维结构，在细胞内执行多种功能。蛋白质被视为万能的纳米机器。

DNA可以编码构建细胞和生物所需的蛋白质，对这一知识的了解为我们在理解细胞和生物系统方面取得巨大进步打开了大门。由于每种蛋白质序列的编码都要用到DNA（每个位置只能存储4个可能值中的一个），而蛋白质一共使用20种氨基酸，所以我们立刻就能推断出编码每种氨基酸（至少）需要三个碱基。事实上，两个核苷酸只能产生16（4^2）种不同的编码组合，无法满足蛋白质使用的20种氨基酸的需要。

有几位科学家通过其他实验发现了遗传密码，遗传密码的作用是指定每组三个DNA碱基如何编码一种特定的氨基酸。1964年，马歇尔·尼伦伯格和菲利普·里德利用克里克等人取得的一项研究成果，即一种氨基酸编码至少需要三个DNA碱基，破译了标准遗传密码的密码子。

在表6－1中，含有“起始”和“终止”的款目表示每个蛋白质序列的开始和结束，尽管我们现在知道，在现实中，从DNA到蛋白质的翻译过程显然要复杂得多，可能会使用其他机制来启动或终止翻译。

我们现在知道，有些生物使用的遗传密码与表6－1所示的情况略有不同。1979年，巴雷尔、班吉尔和德劳因发现人类的线粒体细胞使用另一种遗传密码。此后，人们还发现了其他略有不同的变体，但表6－1中的标准遗传密码仍然具有参考价值，并且在大多数情况下都适用。

随着遗传密码和从DNA到蛋白质再到细胞的基本机制被发现，全面了解生物系统成为可能。随着时间的推移，人们甚至有可能弄清楚DNA的内部结构，即编码特定蛋白质的基因。对细胞行为来说，基因是极其重要的分子机器。不过，大多数基因对生物的可见特征并不会产生直接影响，这与孟德尔的研究正相反。为特定蛋白质编码的基因在生

物细胞中可以发挥多种作用，对DNA结构的了解，以及许多细胞研究技术的发展，促使我们对细胞有了更加深入的理解。如何通过探究基本原理的方式理解细胞的运行机制，是20世纪后40年里最吸引人的研究之一，当然也是21世纪最有趣的挑战之一。

表6–1 标准遗传密码表。每个密码子由三个碱基组成，编码一种氨基酸

第二个字母

第一个字母　　第三个字母

	U	C	A	G	
U	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	U
	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	C
	亮氨酸	丝氨酸	终止	终止	A
	亮氨酸	丝氨酸	终止	色氨酸	G
C	亮氨酸	脯氨酸	组氨酸	精氨酸	U
	亮氨酸	脯氨酸	组氨酸	精氨酸	C
	亮氨酸	脯氨酸	谷氨酰胺	精氨酸	A
	亮氨酸	脯氨酸	谷氨酰胺	精氨酸	G
A	异亮氨酸	苏氨酸	天门冬酰胺	丝氨酸	U
	异亮氨酸	苏氨酸	天门冬酰胺	丝氨酸	C
	异亮氨酸	苏氨酸	赖氨酸	精氨酸	A
	甲硫氨酸，起始	苏氨酸	赖氨酸	精氨酸	G
G	缬氨酸	丙氨酸	天门冬氨酸	甘氨酸	U
	缬氨酸	丙氨酸	天门冬氨酸	甘氨酸	C
	缬氨酸	丙氨酸	谷氨酸	甘氨酸	A
	缬氨酸	丙氨酸	谷氨酸	甘氨酸	G

细胞这种高度复杂的“机器”，是经过数十亿年的进化形成的。所有细胞都拥有共同的祖先，那就是大约40亿年前生命起源时的单细胞生物。所有现存生命形式最后的共同祖先肯定是这些单细胞生物中的一种，我们对它的昵称是卢卡（LUCA），表示“最后一个普遍的共同祖先”（Last Universal Common Ancestor），人们认为卢卡生活在30多亿年前。通过DNA序列分析，我们知道今天存在的所有生物，包括细菌、古核生物和真核生物（图6–2），都是从这个共同祖先进化而来的。

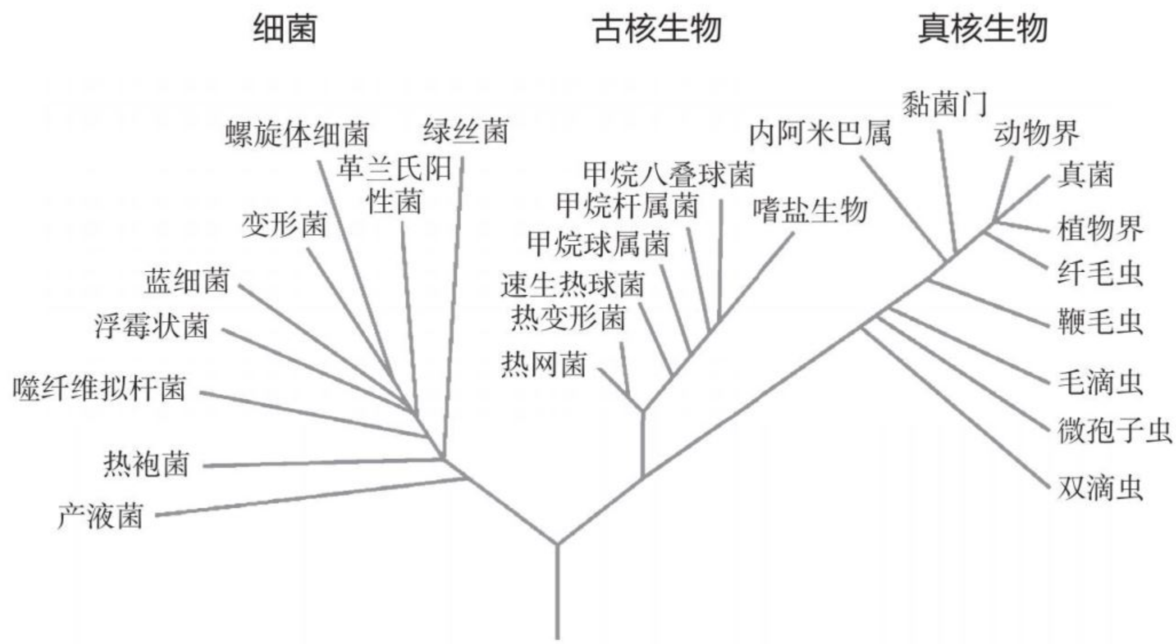


图6-2 基于基因组数据的生命系统发生树
资料来源：卡尔·乌斯，1990。

细菌和古核生物是原核生物（简单的单细胞生物），它们数量众多，能够在所有可能的环境中迅速繁殖。任何多细胞植物和动物，包括人类，都属于真核生物域。也就是说，他们都是由真核细胞构成的。真核细胞是原核细胞经过数十亿年进化的产物，出现时间估计在15亿年前到27亿年前。真核细胞可能起源于更简单的原核细胞：多个原核细胞以共生模式组合在一起，最终变成一个更复杂的细胞。我们无法确定真核细胞是如何出现的，但它们是复杂生命发展过程中的一个重大事件。原核细胞中的大部分遗传物质都包含在一个叫作拟核的不规则形状的区域中，拟核的周围没有核膜。原核生物的基因组通常是一条环状双链DNA，可能有多个拷员。

真核细胞更为复杂，由大量的内膜和结构（叫作细胞器）组成。如果具体描述真核细胞的结构，可以写成几本大部头的书。详细介绍细胞工作原理的优秀参考书有很多，其中一本是杰弗里·库珀和罗伯特·豪斯曼于2000年出版的《细胞》。在这里，我仅以图6－3为例，简要介绍一下真核细胞的一般结构。

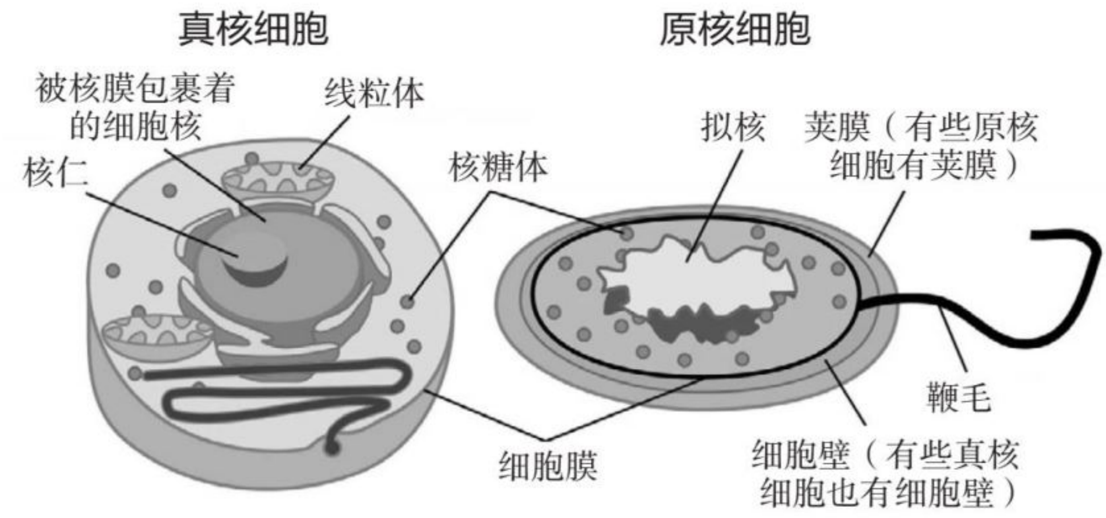


图6－3 真核细胞和原核细胞的结构简图
资料来源：美国国家生物技术信息中心，2007。

真核细胞中最重要的结构是细胞核，细胞核内含有染色体，外面包裹着一层核膜。在细胞核外的细胞质中，有许多起着不同作用的结构。线粒体是参与多种任务的细胞器，它们产生细胞的大部分化学能供应，所以通常被视为细胞的发电机。有人认为线粒体曾经是独立的生物，后来被细胞吸收或以其他方式与细胞结合在一起。线粒体有它们独立的DNA，人类线粒体的DNA大多是由母本传递给后代的。核糖体是一种复杂的分子机器，其质量大约相当于300万个氢原子。核糖体通过一种叫作翻译的过程，利用信使RNA（mRNA）合成蛋白质。大多数真核细胞都有一个由微管和微丝组成的细胞骨架，细胞骨架在确定细胞的组织和形状方面起重要作用。

真核生物的DNA被分成几个线束，即染色体。现在，人们已经比较好地理解了DNA读取和DNA指令被翻译成细胞生物学的中心法则的过程，尽管还有许多方面不太清楚。接下来，我简要介绍一下真核生物中的DNA密码被转化为功能蛋白质的方式。

整个过程从转录开始。在转录期间，细胞核中的一个DNA片断上的信息被转移到新生成的信使RNA上。很多蛋白质都参与了这个过程，包括RNA聚合酶（负责读取DNA中包含的信息）和其他辅助蛋白质，比如

转录因子。转录因子很重要，因为DNA中的所有基因并非都以相同的速度被转录，有些基因可能根本就没被转录。转录速度由转录因子控制，转录因子可以加速、减缓甚至终止特定基因的转录。这是造成细胞分化的机制之一。DNA完全相同的细胞在不同的组织中或在不同的时间点会有不同的表现。在真核细胞中，第一信使RNA的转录常通过剪接来改变。在剪接过程中，一些信使RNA片段被剪掉，剩下的片段被拼接在一起，产生最终的信使RNA。经过剪接处理的信使RNA穿过细胞膜上的小孔离开细胞核，进入细胞质，并在那里与核糖体（细胞中生产蛋白质的机器）结合。

核糖体以三个碱基为一组（即密码子）读取信使RNA，通常从密码子AUG开始。每个密码子都会根据遗传密码被翻译成一个氨基酸。其他分子机器将转运RNA（tRNA）带入核糖体－信使RNA复合体，使信使RNA中的密码子与转运RNA中的反密码子相匹配，从而为蛋白质添加合适的氨基酸。随着越来越多的氨基酸连接到生长链上，蛋白质开始折叠成具有特定功能的三维结构。这种折叠过程会一直持续到生成的蛋白质以折叠的形式从核糖体中被释放出来。折叠过程本身相当复杂，可能需要很多伴侣蛋白质的通力合作。

蛋白质通过原子间相互作用折叠成特定的空间结构，以实现其生物功能。原子间相互作用源于多种可产生引力和斥力的现象。例如，氢原子通过共用电子形成H2分子，并通过共价键（而不是非共价键）保持在一起。非共价键是由带电离子或极性分子之间的静电相互作用形成的。水分子由一个氧原子和两个氢原子构成，由于其内部原子的排列方式，水分子的一部分带有正电荷，而另一部分带有负电荷。这样的分子叫作极性分子，它们通过电磁力与其他极性分子相互作用。

因此，在蛋白质的原子之间，或者在这些原子与环境中的分子（最常见的是水分子）之间会产生一组力，使蛋白质形成某种结构。蛋白质结构是生物学领域的一个重要研究课题。人们利用多种技术，

比如X射线晶体学和磁共振谱学，来解释蛋白质结构。通常认为蛋白质结构有4级，即一级、二级、三级和四级结构。

蛋白质的一级结构是氨基酸链中的氨基酸线性序列，它取决于编码蛋白质的DNA序列。二级结构是指蛋白质某些部分的有规律的亚结构，它使氨基酸链形成特定的几何排布，主要有α螺旋和β片层两种类型。三级结构是指蛋白质分子的三维排布，α螺旋、β片层与氨基酸链的其他部分一起被折叠成一个致密结构，使其在与环境（通常是水或其他蛋白质）接触时分子的总能量最小。折叠过程是由多种原子间相互作用驱动的，包括把疏水残基安放在远离水的地方，但结构主要由特定的原子间相互作用（比如非共价键、氢键和侧链的致密堆积）来决定。四级结构是指不同蛋白质之间的相互作用，包括决定不同蛋白质的相互作用和相互停靠方式。蛋白质复合体的种类繁多，既有非常简单的双蛋白排布（二聚体），也有由大量亚基构成的复合体。图6－4所示的核糖体复合体由几十种蛋白质构成。

人们提出了很多利用计算机确定蛋白质的二级、三级和四级结构的计算方法。确定蛋白质结构的能力，对我们理解诸多生物现象和设计新药来说至关重要。今天，通过思考一种蛋白质和其他蛋白质的相似性，以及每种蛋白质结构的总能量，我们有可能利用计算机预测出这种蛋白质的空间结构。为测试这些方法并比较它们的准确性，人们组织了一些竞赛。其中，蛋白质结构预测大赛自1994年起每两年举办一届。

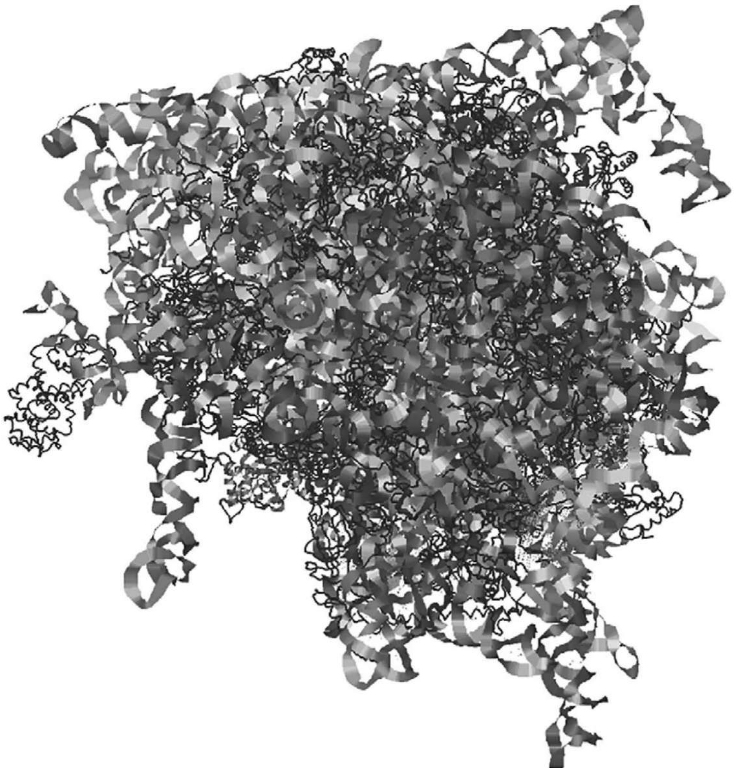


图6－4 人类80S核糖体的结构
资料来源：蛋白质数据库。

不过，如果待分析的蛋白质与其他已知蛋白质不相似（至少部分不相似），这个问题基本上就无法解决，因为在这种情况下，寻找正确的空间结构将面临计算困难。图6－5（a）显示的是小蛋白质1R69p的三级结构示意图，它只包含5个短的α螺旋。即使是这样一个小蛋白，从头开始（也就是说，仅凭它的一级序列，而不利用它和其他已知蛋白质的相似性）确定它的空间结构，也不是一件易事。图6－5（b）展示的是利用只能得出近似结果的计算方法，从头开始确定该蛋白质的三级结构得到的结果。

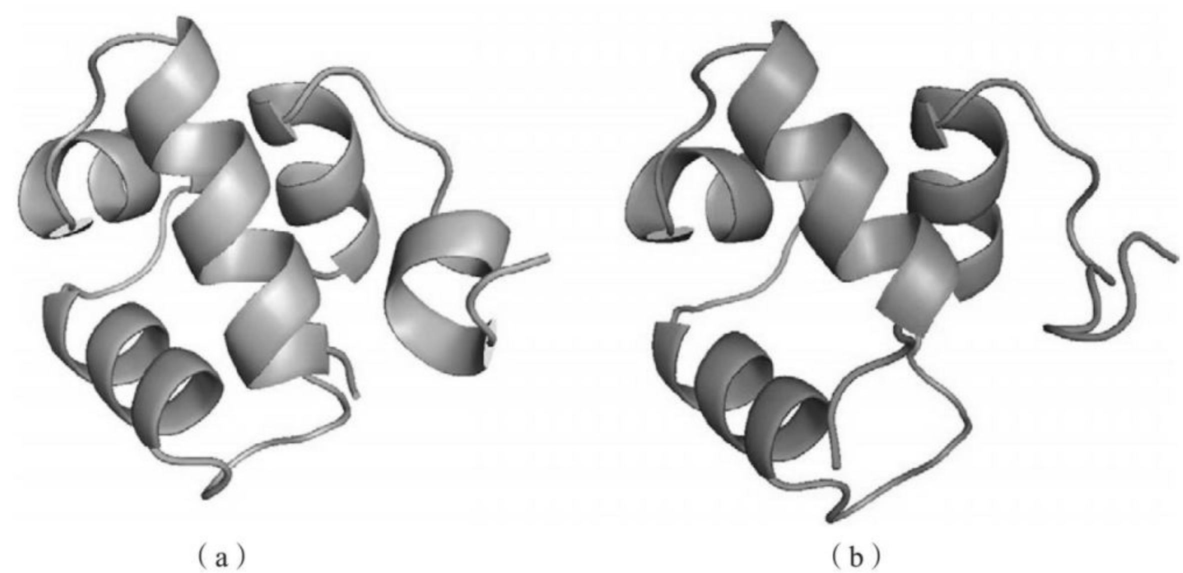


图6－5 蛋白质1R69p的三级结构

在任何时刻，除了转录和翻译之外，细胞内部还运行着其他许多过程。其中有些人们已经十分了解，而对其他过程却还一无所知。不同的生化过程调节着营养物质和废物进出细胞的速度，控制着繁殖周期，并确保有足够的能量供细胞运行。由于细胞内的许多过程都需要能量，所以大部分细胞活动的目的都是获取能量。细胞利用储存在ATP（5’－三磷酸腺苷）中的化学能——大多是由线粒体产生的——为许多消耗能量但又必须发生的化学反应提供能量。在几乎所有细胞内都很常见的糖酵解过程中，葡萄糖转化为丙酮酸时会产生两个ATP分子的净增益。这就是几乎所有活细胞都需要葡萄糖（糖的一种）来为它们的生物过程提供能量的原因。人们已经非常了解这一过程和许多其他复杂的代谢过程，并且翔实地记录了转化路径。

大多数真核细胞的大小为2~100微米。真核细胞看似很小，但我们要知道，碳原子的半径约为70皮米，核糖体的直径约为0.3纳米。如果把一个直径为25微米的真核细胞像吹气球一样吹大，使其直径变成1英里，碳原子就只有蓝莓那么大，核糖体的直径也只有6英尺。注这意味着真核细胞内部有足够的空间容纳几百个甚至几千个核糖体和几百万个其他细胞机器。尽管人们投入了巨大的精力试图了解细胞的行

为，但除了比较简单的过程之外，我们对细胞机制的理解仍然是零碎和不完整的。

生物学的终极目标不仅是了解特定细胞过程的行为，还要了解器官甚至整个生物的行为。现在，为了实现这个目标，人们利用计算机和算法处理实验数据和建模。在描述如何使用计算机和算法研究生物系统并为其建模之前，我们有必要先讨论为什么身体和大脑对真核细胞的生存来说必不可少。

身体和大脑

多细胞生物曾多次独立崛起，第一次可能是在10亿多年前。大约6亿年前，在寒武纪大爆发（地球环境发生剧烈变化的一个时期）之前和期间，最复杂的多细胞生物——动物——的数量骤增。

造成多细胞生物崛起的原因，可能是细胞在分裂过程中无法分离，也可能是细胞偶然地粘在一起。如果自然选择偏好这种新的生命形式，它就会生存下来并不断发展。多细胞生物有许多可能的优势。由于较大的生物总有生态位，所以进化可能偏爱较大的物种，而且这些物种必然是多细胞生物。随着生物变得越来越大，为了满足维持更大体型所需的更大复杂性，它们必须发展出专门的细胞类型。

随着时间的推移，进化的盲算过程使细胞类型发生分化，复杂的多细胞生物也越来越多。分化的细胞类型在复杂动物身上形成了不同的器官，最终还发展出一个专门处理信息的器官——大脑。人们逐渐发现，在多细胞生物中观察到的复杂性，大多不是源于细胞机制层面的重大创新，而主要源于基因调控机制，其中一些调控机制在转录后发生作用，导致了细胞的分化和专门化。对有些人来说，当想到他们的身体是由相当普通的细胞组成的，而且这些细胞与许多其他动物的细胞非常相似，甚至与面包酵母菌等单细胞生物也有些许相似之处时，他们可能会有点儿沮丧。事实上，人类的基因数量与小鼠及大鼠大致相同，更比水稻和其他一些物种少，所以不太可能有一组专门用于制造人脑的基因。人类的基因不仅在数量上和老鼠差不多，在类型上也非常相似。即使是不起眼的果蝇也和我们有许多相同的基因，果蝇和人类用来确定身体结构的基因非常相似，都是从一个遥远的共同祖先（生活在6亿年前的某种扁形虫）那里继承来的。人类大脑的编码

基因可能与猴子大脑的编码基因基本相同，只是在基因调控上有一些小的不同之处，结果导致了截然不同的大脑特征。

从复制因子的角度看，身体只是确保复制成功的一种便利机制。独自在海洋中游动的单个细胞，非常容易受到各种各样的攻击。多细胞生物拥有高度专门化的细胞，这些细胞共享相同的遗传物质。通过这种便利的方式，可以确保多个细胞为成功复制相同的基因而共同努力。随着时间的推移，身体变得高度专门化，仅凭蛮力不再能确保生存。因此，复制因子想出了一个新办法。它们通过变异，编码并产生了信息处理细胞，用它们来感知环境和预测行动的可能结果。这些细胞可以远距离传递信息，还可以组合不同来源的数据，以辅助决策。其中一些细胞开始利用电信号传递信息，电信号的传播速度比生化信号快，使用形式也更加灵活。

人们在一种叫作领鞭毛虫的单细胞生物体内，发现了传递电信号和通过化学信号调节电信号所需的多种机制。领鞭毛虫是多细胞动物现存亲缘关系最近的单细胞物种，属于单细胞鞭毛真核生物，细胞体呈卵形或球形，有单鞭毛。通过鞭毛的运动，领鞭毛虫可以在水中自由游动。

道金斯认为，领鞭毛虫是在大约9亿年前从产生人类的那条进化路线中分离出来的，因此，如果它们与人类同源（这个假设有可能成立），那么传递电信号的机制出现的时间不会晚于上述时间点。大约在同一时间，细胞发展出通过电脉冲及当时仍在使用的化学信号同其他细胞交流的能力。

随着时间的推移，神经元——能够产生电冲动的细胞——进化出轴突，它可以远距离传输电信号。神经元通过释放化学物质，将这些信号传递至其他神经元，电信号的传递只能在突触（神经元与其他神经元或肌肉细胞相互接触的地方）上进行。由于细胞能用电信号（而不只是化学信号）传递信息，所以拥有这些细胞的早期生物取得了一

种竞争优势。它们可以更迅速地应对威胁，更高效地寻找食物。对早期海洋中的生存斗争来说，即使少数几个神经元也大有帮助。

能够区分光明和黑暗，或者听到响声，这些能力都是有益的。不过，感知外部世界和在采取特定行动之前预测结果的能力，对生存来说用处更大。有了更复杂的大脑后，生物只需要估计每一个行动的成功概率，就可以从几个可能的行动中选出最好的一个，而无须冒险将所有行动都尝试一遍。基因结构不同，神经元的连接就会不同，进化形成的组合信息的方式也会不同——一些比另一些更有用，从而使动物采取安全或不那么安全的行为。最有用的神经元结构激增，因为它们的主人由此取得了竞争优势。经过几百万代的进化，这种信息处理装置变得越来越精妙复杂。一开始是简单的信息处理设备，可以分辨光明与黑暗，听出有没有声音；后来进化成复杂的信息处理电路，可以识别远处的敌人或猎物，并根据情况控制主人的行为。

最早的神经元可能分散在早期动物的身体各处。但随着神经元复合体变得越来越复杂，神经元开始聚集在一起，形成了早期的中枢神经系统。神经元彼此接近意味着可以快速处理更复杂的信息，尤其是在最需要这种能力的地方，即嘴巴和感光装置（后来变成了眼睛）附近。虽然我们还不清楚原始大脑产生于何时以及经过了多少次发育，但脊椎动物的祖先（一些类似鱼的古代生物）拥有类似大脑的结构。随着越来越复杂的信息处理设备的发展，它们展开的“军备竞赛”产生了今天人类大脑中的许多结构，包括基底核（控制运动模式）和视顶盖（参与追踪运动物体）。最终，作为进化压力的直接结果，这种“军备竞赛”促使大脑变得越来越聪明。达尔文在《物种起源》中写道，“智力水平取决于物种完成生存所需活动的效率。”

大约2亿年前，早期脊椎动物的后代迁移到陆地上，并继续进化。其中一个进化分支最终产生了哺乳动物，早期哺乳动物的大脑有一个体积较小的新皮质，它是负责哺乳动物的行为复杂性和灵活性的脑区。在与恐龙竞争的过程中，哺乳动物的大脑体积不断增大，这改善

了它们的嗅觉、视觉和触觉。恐龙灭绝后，一些幸存下来的哺乳动物爬上了树，成为灵长类动物的祖先。良好的视力不仅在复杂的环境中用处颇大，还促使大脑新皮质的视觉部分变大。与此同时，灵长类动物也发展出更大、更复杂的大脑，可以整合、处理它们收到的信息，并根据这些信息规划和控制它们的行动。

因此，生活在1 400万年前的非洲类人猿聪明机敏，但可能并不比它们的非人类后代——猩猩、大猩猩和黑猩猩聪明多少。与之相比，人类的进化方式完全不同，速度也极快。大约280万年前，随着能人在非洲出现，人类大脑开始快速发育，但我们并不清楚其中的原因。很可能是某种突变或者是一系列小突变，推动了大脑的发育，使人属的各个物种足够聪明，能够使用技术和发明语言。从早期物种到人属成员，其间实现的最重要发展是，脑容量由早期物种的约450立方厘米增加到能人的600立方厘米。

在人属中，当能人于大约60万年前进化成海德堡人时，脑容量又增加了一倍，达到与智人大脑差不多的大小——大约1 200立方厘米。其间的数千年时间里，人属的多个物种共同生活在地球上，包括在大约4万年前灭绝的尼安德特人，以及存续到大约1.2万年前的佛罗里斯人。

1.2万年以来，人类一直是生活在地球上的人属的唯一代表——自从近300万年前人属出现以来，这种情况绝无仅有——这可能是一个历史巧合吧。产生这种独一无二现象的主要原因，可能是智人独有的组织大型社会和改变环境的能力。

上述的所有发展共同打造出现代的人类大脑，这是我们已知的最复杂的信息处理机器。大约20万年前，随着智人的崛起，现代人类大脑在非洲出现了。了解这样一种大脑的运行方式，是科学界有史以来面临的最有趣的挑战之一。但在应对这一挑战之前，我们有必要采取迂回策略，先研究一下计算机在我们理解复杂生物系统过程中发挥的作用，以及如何利用计算机为人脑建模。

1. 1微米= 10^{-6}米，1皮米= 10^{-12}米，1纳米= 10^{-9}米，1英尺≈ 0.3米。——译者注

第7章

生物学与计算机的美妙邂逅

如果说20世纪是物理学的世纪，那么21世纪很可能是生物学的世纪。有人或许会辩称，生物学是一门历史悠久且丰富的科学，而21世纪将是多种技术百花齐放的世纪。多种技术齐头并进自然有其好处，不过，生物学必然会成为21世纪的决定性技术之一，就像20世纪的物理学一样，生物学在21世纪的头几年也取得了令人兴奋的进展。正如物理学在1905年因爱因斯坦发表的那几篇影响深远的论文而发生改变一样，生物学在2001年也因为两个竞争团队发表的两幅关于人类基因组的全部DNA序列的草图而改变。

基因组测序

早在20世纪50年代早期沃森和克里克完成他们的开创性研究之时，人们就已经了解了DNA的结构及其在遗传学中的作用，但在此后大约20年的时间里，人们却一直无法解码任何生物基因组中的碱基序列。然而，人们从一开始就很清楚，解码人类基因组将给我们带来新方法，可用于诊断和治疗遗传性疾病，以及解决由个人基因组成决定的许多其他问题。

人类基因组测序是在DNA结构被发现后的几十年里开展的一项重要研究活动。1990年，对测定人类基因组序列的兴趣最终促成了人类基因组计划的诞生。该计划由詹姆斯·沃森领导，它的主要目标是测定构成人类DNA的化学碱基对序列，识别并绘制人类基因组中的数万个基因。最初，这项计划预计耗资30亿美元，耗时15年，由美国、中国、法国、德国、日本和英国的遗传学家组成的国际团队来承担测序工作。

1998年，克雷格·温特和他的塞雷拉基因组公司发起了一个类似的但由私人资助的计划，希望以更快的速度和以人类基因组计划的一小部分成本，完成基因组测序工作。两个计划使用的方法不同，但在2001年几乎同时达成了获取人类基因组的第一张草图的目标。

支撑人类基因组测序活动的测序技术，并未在2001年停下其以指数速度发展的势头，这是包括爆炸性发展的数字电路设计在内的其他任何领域都无法相提并论的。测序技术的迅猛发展促使基因组测序项目数量激增，21世纪初，完成全基因组测序的物种数量急剧增加。

测序技术的发展速度甚至超过了微电子技术。在人类基因组计划启动之时，DNA测序的平均成本是每个碱基对10美元。2000年，该成本降为原来的1%，此后也在不断下降（图7－1）。2015年，每个碱基对的测序原始成本约为10^{-6}美分，并且继续下降，只不过降速比以前慢了。人类基因组计划的预算为30亿美元（每个碱基对约为1美元）。2007年，沃森在他自己的基因组测序中使用了一项新技术，共计花费200万美元。这个成本看似高昂，但它还不到人类基因组计划预算的千分之一。在我撰写本书时，人类基因组测序的成本正在接近1 000美元这个人们追求已久的目标。在15年里，每个DNA碱基对的测序成本降至原来的20万分之一，年下降速度超过50%，大大超出了摩尔定律预测的成本下降速度。根据摩尔定律，每两年成本就会减半。

尽管人们对用于测定DNA特定链中的特定碱基序列的方法越来越感兴趣，但在发现DNA之后的几年里，测序技术的起步速度非常慢。从发现DNA双螺旋结构到发明第一种测定DNA碱基序列的实用方法，历经了20年时间。由于不可能直接观察DNA链上的碱基序列，所以必须诉诸于间接的方法。在早期使用的方法中，最实用的一种是双脱氧链终止法，我们现在称之为桑格－库森法。这种方法依据的基本原理是，先利用特定的分子使脱氧核苷酸（DNA的基本成分）溶液中的DNA链终止生长，然后根据大小将DNA序列分解开，使读取碱基对序列成为可能。该方法先通过一个叫作变性的过程获得单链DNA，变性就是用加热的方式将DNA的双螺旋拆分为两条单链。然后，以单链DNA为模板，通过4个独立的物理反应，向其中添加A、C、T、G这4种标准的DNA脱氧核苷酸，从而创建出新的双链DNA。每个反应不仅会使用标准的脱氧核苷酸，还会使用一种经过修饰的脱氧核苷酸（即双脱氧核苷酸）溶液，后者有终止双链DNA生长的属性。每种链终止子都会附着在单链DNA的某个特定碱基（A、C、T或G）上，阻止DNA双链进一步生长。之后，让DNA单链流经凝胶，因为不同大小的DNA片段在凝胶中的流动速度不同，所以它们被分解开来。最后，根据DNA片段在凝胶中的位置，我们就能读出原始DNA片段中的碱基序列。

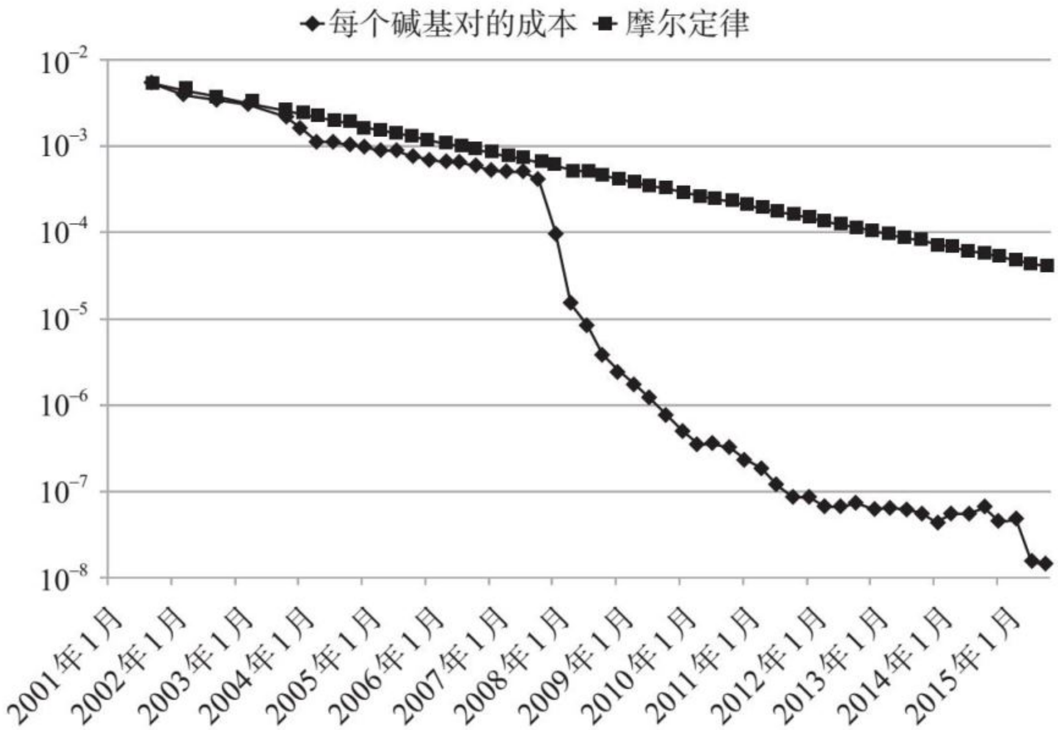


图7－1 依据美国国家人类基因组研究所的汇编数据绘制的测序成本与摩尔定律的对比图（单位：美元/DNA碱基对）

桑格－库森法使得研究人员可以处理小的DNA序列，因此它首先应用于小生物子系统的研究。对特定基因或其他小DNA序列的深入研究，增加了人们对细胞现象的理解。

在长达几十年的时间里，已有的技术一直无法完成任何生物的全基因组测序。随着技术的发展，人们可以测定越来越长的DNA序列，并在1972年第一次完成了全基因组测序，在1976年第一次完成了生物的全基因组测序（包含3 569个碱基的噬菌体MS2的DNA序列）。但是，要完成复杂生物的全基因组测序，还有一段漫长的路要走（复杂生物的基因组中有数十亿个碱基）。

1985年，人类基因组测序的第一个现实方案出现了。但直到1988年，人类基因组计划才明确表示，这样一个雄心勃勃的目标是可以实现的。通过选择复杂程度递增的模式生物，人类基因组测序必需的技术和基因组学的研究，都取得了渐进式发展。被选中进行测序的模式生物包括：1995年，流感嗜血杆菌（1.8兆碱基）；1996年，大肠杆菌

（5兆碱基）和酿酒酵母菌（面包酵母菌，12兆碱基）；1998年，秀丽隐杆线虫（线虫的一种，100兆碱基）；2000年，拟南芥（小型开花植物，125兆碱基）和黑腹果蝇（200兆碱基）。2001年，上述研究的最后几步派生出两篇关于人类基因组的论文，分别发表在《科学》与《自然》杂志上。这两项研究使用的方法不同，但并非毫无关联。

人类基因组计划使用的测序方法与最初用来测定酿酒酵母菌和秀丽隐杆线虫基因组序列的方法相同。在项目启动之时，可用的技术是BAC（细菌人工染色体）测序法。BAC是人工制造的DNA序列，用于转化和克隆细菌（通常是大肠杆菌）的DNA序列。BAC的大小通常为150~350千碱基。

该测序法的第一步是创建基因组的粗略物理图谱。构建这张图谱需要将染色体切分成若干大块，先确定这些大的DNA片段的顺序，再测定单个DNA片段的序列。为了达到这个目的，可以随机选择几个位置，将基因组的几个拷贝切割成长度约为150千碱基的片段，并将每个片段插入BAC。包含人类基因组片段在内的BAC集合叫作BAC文库。利用“指纹识别”，为这些大块中的DNA片段分别赋予一个独特的识别标签，据此确定这些片段的顺序。具体来说，就是利用单酶切割每个BAC片段，在重叠的片段中找到共同的序列，从而确定每个BAC在染色体中的位置。大约每隔10万个碱基，BAC就会发生重叠，并带有这些标记，因此，我们利用这些标记就可以创建每条染色体的物理图谱。接下来，每个BAC被随机切分成更小的块（长度为1~2千碱基）并排序，通常是从片段的每一端切下大约500个碱基对。随后，这一过程产生的数百万个序列被组装到一起（通过一种寻找重叠序列的算法），并被放到基因组的恰当位置上（根据生成的物理图谱）。

然而，塞雷拉公司提出并采用的“霰弹枪”测序法，则不需要创建物理图谱。基因组的拷贝被随机切分成长度为2~10千碱基的片段，并且都被插到质粒（细菌中可以复制的DNA片段）中。然后，对质粒文

库进行测序，利用每一端的大约500个碱基对，得到几百万个片段的序列。组装所有这些序列时，我们需要使用复杂的计算机算法。

测序技术的进一步发展催生了今天的许多技术，它们大多基于不同的生化原理。现在的测序仪每次运行都能产生几百万个DNA读序。一次运行需要花两到三天时间，生成的数据量接近1太字节（10^{12}字节），比整个人类基因组计划生成的数据总量还要多。

每个读序都包含一小段被分析的DNA序列，根据测序技术的不同，这些片段的大小从几十个到几百个碱基对不等。目前，市场上有一些相互竞争的技术，其中最常见的是焦磷酸测序、连接法测序、合成法测序和离子激流测序。与下一代测序仪相比，桑格–库森法在每个位置上提供的读序更长、错误更少，但吞吐量却低得多，因此使用成本非常高昂。

考虑到DNA序列的数量和长度，计算机科学家认识到根据读序测定基因组，甚至要从大量数据中找出相似序列并不是一件容易的事。于是，他们开始尝试用算法处理生物序列。

因为所有生物都有共同的祖先，所以不同生物的DNA有很多相似之处。找到这些相似之处将大有帮助，因为一个特定基因在一种生物中发挥的作用很可能与一个相似基因在另一种生物中发挥的作用类似。用于处理生物序列（特别是DNA序列）的有效算法的开发，催生了一个新的科学领域——生物信息学。对人类基因组测序来说，可以处理非常长的符号序列（即字符串）的算法尤其有用。虽然处理小字符串对计算机而言相对容易，但当字符串像DNA序列一样包含几百万个符号时，这项任务的难度就会大大增加。于是，如何有效处理字符串成为计算机科学中最重要的领域之一，处理字符串的有效算法在生物信息学和网络技术中都占据着关键地位。

寻找完全或近似匹配的生物序列的BLAST程序，是一个非常成功的生物信息学分析工具。目前，BLAST仍被广泛用于数据库检索，旨在找

出与DNA序列或人们感兴趣的蛋白质序列高度匹配的序列。生物信息学领域的研究人员开发了可以处理DNA字符串的许多其他算法，其中一些专门用于将测序器生成的几百万个片段组装成一个个染色体序列。

尽管细菌人工染色体测序法和霰弹枪测序法看起来相似，都是将小的序列片段组装成染色体的完整序列，但它们之间其实有着天壤之别。在细菌人工染色体测序法中，物理图谱是已知的，而且易于找到与片段相对应的序列，因为我们大致知道片段的位置。这就好像玩拼图游戏时，如果根据颜色我们知道了各个图块的大致位置，这个游戏玩起来就不会太难。但是，对人类基因组来说，创建BAC文库并推导出物理图谱是一个漫长又复杂的过程，需要花费许多年的时间来完善和实施。而在霰弹枪测序法中，这个过程快得多，难度也小得多，因为长DNA序列直接被切分成小片段，所以无须构建物理图谱。此外，在不知道各个片段在基因组中的大概位置的情况下组装几百万个片段，其计算难度要大得多，就好比在玩拼图游戏时所有图块的颜色都一样。只有在算法和计算机技术取得重大进步之后，研究人员才有可能开发出可以高效完成这项工作的程序。

有一种算法能够依据DNA读序构建DNA序列，它建立在根据读序构建图形（我们称之为德布鲁因图）的基础之上。在德布鲁因图中，每个节点对应一个包含k个碱基的长度固定的特定片段；如果在某个包含k+1个碱基的读序中，有两个包含k个碱基的片段相互重叠，图中的两个节点就是相连的。

例如，假设有以下四碱基读序：TAGG，AGGC，AGGA，GGAA，GAAT，GCGA，GGCG，CGAG，AATA，GAGG，ATAG，TAGT。通过在多个位置处进行切分，能够产生所有这些读序的最短DNA序列是什么？绘制如图7－2所示的图形就可以解答这个问题，图中每个节点对应一个碱基三联体，每条边对应一个读序。因此，找出包含所有这些读序的最小DNA序列，就相当于在这个有向图中找出一条欧拉路径——从图中的一个节点出发并恰好经过每条边一次的路径。正如我们在第4章看到的那

样，当且仅当恰好有两个顶点的度为奇数时，无向图中才会存在这样一条路径。在这幅有向图中，我们必须对欧拉定理稍加修改，因为我们只能沿一个方向经过所有边。如果除起点和终点以外的所有顶点的入度和出度相同，而且起点（终点）的出度等于终点（起点）的入度加1，该有向图中就存在欧拉路径。在图7－2中，除了顶点TAG和AGT（它们分别满足起点和终点的条件）以外，所有顶点的入度都等于出度。因此，欧拉路径从节点TAG开始，经过节点AGG和图上部环路中的所有节点后返回到节点AGG，并在依次经过图下部环路中的所有节点后至节点AGT结束。这条路径给出的序列是TAGGCGAGGAATAGT，也就是包含所有给定读序的最短DNA序列。

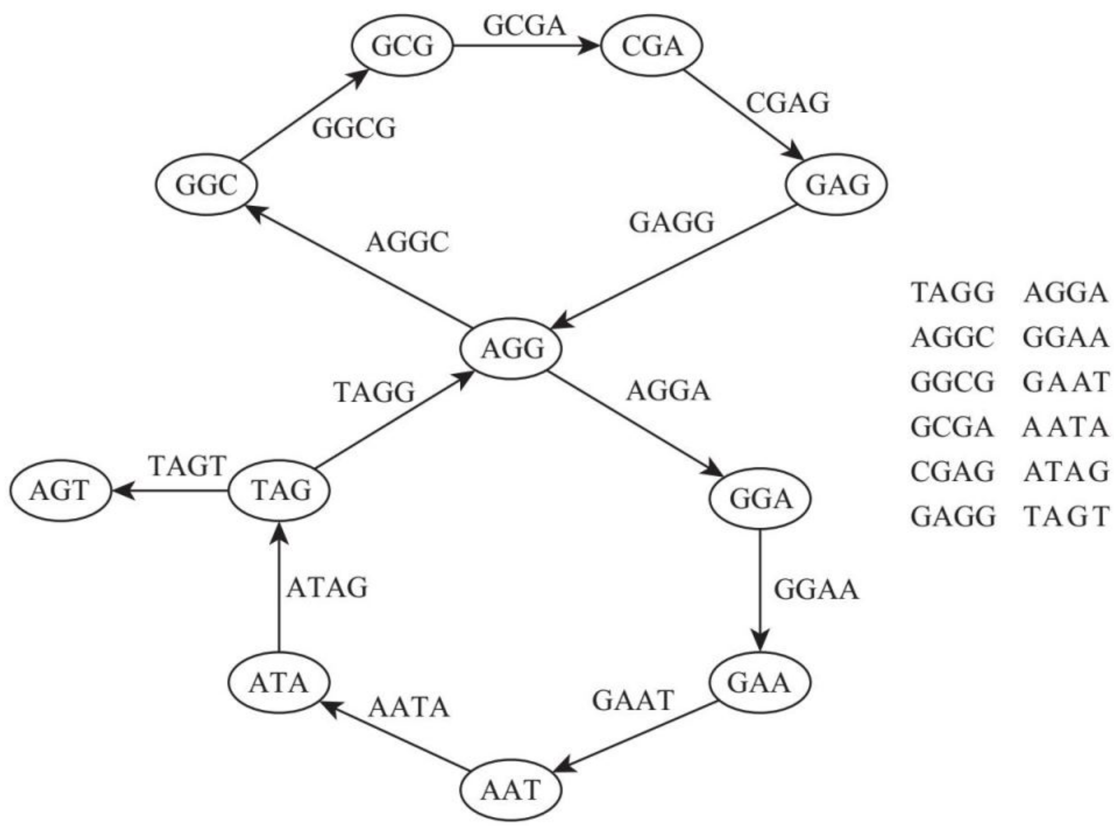


图7－2 利用欧拉路径和根据读序完成的DNA测序。右侧读序用于标记图中的各条边，如果用来标记连接两个节点的边的读序中包含这两个节点的标签，就表明这两个节点是相连的

这个问题也可以改为这样一种表述方式：在由顶点表示读序和由边表示读序之间的重叠部分的图形中，找出哈密顿路径。人类基因组计划使用的正是这种方法，由于哈密顿路径问题是一个NP完全问题，

所以这种方法的计算量更大，效率更低。实际上，由于序列错误、序列缺失、DNA区域重复率高和其他一些因素，使用德布鲁因图组装DNA序列比图7－2展示的要复杂得多。

几乎以平局告终的人类基因组测序竞赛清楚地表明，计算机和算法在生物研究领域可以发挥非常重要的作用。事实上，如果没有这些算法，人类基因组的组装就不可能实现。所以，算法是为实现这一目的而开发的最重要的技术之一。生物信息学领域伴随这些项目而诞生，从那时起，它就吸引了世界各地的诸多研究团队为之奋斗。

今天，成千上万的研究人员和专业人士从事生物信息学工作，利用计算机解决生物学和医学问题。现在，生物信息学（又被称为计算生物学）是一个包含诸多子领域的广阔领域。事实上，处理DNA序列只是计算机执行的众多常规任务之一，它将辅助我们走过漫漫征程，使我们对生物系统产生更加全面的理解。生物信息学领域的研究人员和专业人士多才多艺，解决了各种各样的难题，包括为多种类型的生物网络建模并进行动态模拟。这些网络——用于研究生物系统行为的抽象模型——是我们探索和理解大自然的基本工具。

生物网络

多细胞生物是由数以亿计的细胞组成的。人体包含超过30万亿个细胞，这个数字大到超乎我们的想象。

虽然DNA含有生物的遗传信息，但活细胞内还有很多其他成分。借助近年来开发的技术，研究人员可以获得关于细胞内的各种化学物质浓度的海量数据。对希望了解特定细胞行为和一般生物系统的研究人员来说，这些测量结果非常重要。基因组学取得的进展与其他生物技术领域的进步并驾齐驱，测序中心和生物研究机构收集的数据量也正在以越来越快的速度增长。

为了理解这些数不胜数的数据，生物过程的计算机模型被用来研究提供这些数据的生物系统。生物模型种类繁多，其中应用最广泛、最有效的模型是生物网络。一般来说，把复杂的生物系统看作一组网络（系统的各构件相互作用），是一种有益的做法。生物网络通常可以用图表示，其中的每个节点代表网络的一个构件，节点之间的边表示构件之间的相互作用。

人们利用多种类型的生物网络为生物系统建模，而且每类网络都有多个子类型。在这里，我将以少数有代表性的生物网络类型为例，说明如何利用生物网络模型来研究、模拟和理解生物系统。

一般来说，生物网络的节点不一定与现实的物理位置相对应。在这种情况下，生物网络代表着一种有用的抽象物，但生物网络与路网或电网不同，其节点和物理位置之间并不存在一一对应关系。我接下来要讨论4种类型的生物网络，其中只在神经元网络中节点和特定的空间位置之间存在直接的对应关系，但也不是完美的对应关系。

这4种生物网络是：代谢网络，蛋白质相互作用网络，基因调控网络和神经元网络。我将使用神经元网络这个术语来指代为生物神经元建模的网络，以便将它们与神经网络区分开。尽管其他类型的生物网络也被用来给生物系统建模，但研究细胞、系统乃至整个生物行为的模型大多是由这4种类型的网络构建的。

代谢网络的作用是为活细胞的化合物之间的相互作用建模，这些化合物通过生化反应相互转化并建立联系。人们通常利用代谢网络为细胞内的各种化合物浓度的动态变化建模。这些网络不仅可以描述代谢的化学反应结构和动态变化（也称代谢途径），还可以描述化合物间相互作用的调节机制。这些反应都会受到酶的催化，酶是选择性催化剂，可以加速大多数生物过程（从营养素的消化到DNA的合成）中的代谢反应。虽然酶有多种类型，但大多数酶都是蛋白质，并在DNA中编码。如果一种生物的全基因组是已知的，我们就有可能部分或全部重建细胞内的生化反应网络，并推导出可以描述每种化合物浓度的动态变化的数学表达式。

蛋白质相互作用网络极其广泛地应用于为细胞或细胞系统的动态变化建模。由基因编码的蛋白质通过相互作用，执行细胞的绝大部分功能，细胞中的许多基本过程都是由若干蛋白质组分构建的分子机器完成的。蛋白质之间的相互作用是生化反应的结果，但主要是因为静电力。事实上，蛋白质的四级结构就是由若干折叠蛋白质形成多蛋白质复合体的排布。蛋白质相互作用网络的相关研究，是我们理解活细胞内部的多个过程，以及生物体内不同细胞之间的相互作用的基础。

基因调控网络的作用是为导致不同环境条件下的基因活性差异的过程建模。尽管多细胞生物的所有细胞（直接参与有性生殖的配子除外）都拥有相同的DNA，但复杂生物体内有各种各样的细胞，一个细胞类型还有多种不同的状态。在成年生物体内，细胞分化的过程通常从干细胞开始，将专门化程度不高的细胞类型转变成更加专门化的细胞。细胞分化使得细胞的大小、形状和功能各不相同，这主要受基因

表达变化的控制。不同类型细胞（甚至是类型相同但状态不同的细胞）的基因表达不同，因此，尽管不同的细胞拥有相同的基因组，但它们的特征可能迥然不同。即使一个细胞也可能有多种不同的状态，这取决于特定时间的基因表达。基因活性受到若干转录前和转录后调控机制的控制，最简单的调控机制与转录的启动有关。只在一些蛋白质（被称作转录因子）附着到基因编码区启动端的邻近区域（启动子区域）后，基因才能被转录（产生信使RNA，然后被翻译成蛋白质）。如果出于某种原因，转录因子没有附着到启动子区域，基因就不会被转录，也不会产生蛋白质。为这种机制建模的基因调控网络，实际上就是DNA－蛋白质相互作用网络。该网络中的节点要么是转录因子（蛋白质），要么是基因（DNA片断）。基因调控网络可用于细胞行为的一般性研究，也可用于细胞分化的具体研究。了解基因调控网络的行为和动态变化，是理解复杂生物系统行为的重要一步。

神经元网络是一种广泛应用于大脑过程研究的细胞相互作用网络。虽然人们经常使用不同类型的细胞相互作用网络，但神经元相互作用网络（神经元网络）最值得我们关注。复杂生物的神经系统由相互连接的神经细胞构成，其他类型的细胞则起支持作用。为这种神经元网络建模，是深入理解神经系统行为的关键一步。神经元网络中的节点对应神经细胞，边对应神经元之间的连接。每个神经元的活动水平和神经元之间的（兴奋性或抑制性）连接，决定了神经系统中神经元的兴奋模式，并与特定的大脑行为相对应。在人类和人类的近亲灵长类动物的中枢神经系统中，这些行为模式产生了智能行为和可能的意识行为。因此，神经元网络研究对理解大脑行为来说极其重要。

正如我们在第5章看到的那样，人们利用简化的神经元数学模型创建人工神经网络。前面说过，人工神经网络可以表现出神经元网络的一些特征，尽管它们不是生物系统的真实模型。这些人工神经网络被广泛地用作脑区模型，以及执行要求不断适应和学习的复杂计算任务的机制。但总的说来，人工神经网络的作用并不是为大脑中的真实细

胞行为建模。用更复杂的神经元模型来模拟活体大脑中的细胞行为，是第8章要讨论的内容。

生物系统建模与仿真

精确的生物系统模型的开发，为我们在硅片上或使用计算机模拟真实生物系统的行为创造了条件。系统动态行为的模拟，在许多其他科学领域都很常见。物理学家模拟原子、分子、机械体、行星、恒星、恒星系统乃至整个星系的行为。工程师常通过模拟来预测电子电路、发动机、化学反应器和许多其他系统的行为。天气预报员模拟大气、海洋和陆地的动态变化，从而以越来越高的准确度预测第二天的天气。经济学家模拟经济系统的行为，并在一定程度上模拟整个经济体的行为。

在计算机科学中，“模拟”（simulate）一词有一种非常特殊的含义，它通常指计算机之间的模拟。在理论计算机科学中，如果一个计算引擎可以模拟另一个，在某种意义上它们就是等价的。正如我在第4章中指出的那样，在这种情况下，我会用“仿真”（emulate）一词来代替“模拟”。我们已经讨论过可以仿真其他所有图灵机的通用图灵机，也知道通用图灵机可以仿真现有的计算机，即使它的速度很慢。实际上，“虚拟化”（virtualization）已经成为一个常见概念，它是指一台操作系统与程序都完备的计算机被另一台计算机上运行的程序仿真，而另一台计算机本身可能只是一个在其他计算机上运行的仿真程序。当然，这种回归会在某个时刻结束，而且必然有一台计算机是真实存在的，但这种计算的虚拟化理论上不会受到任何限制。

想要模拟系统的行为，需要有一种抽象的计算机制，无论它是模拟的还是数字的。虽然仿真模拟器在过去已被投入应用，在一定范围内还将继续使用，但数字计算机的发展使我们灵活有效地模拟非常复

杂的系统成为可能。模拟复杂系统的能力之所以重要，有多个原因，其中最重要的是：调整时标的能力，观察目标变量的能力，理解系统行为的能力，以及为并不真实存在或无法访问的系统建模的能力。

与真实物理系统的行为有关的时标，可能会导致我们无法观察到这些系统中的事件，要么因为事件发生得太缓慢（比如两个星系的碰撞），要么因为事件发生得太快（比如，分子的动态变化）。模拟可以通过压缩或延长时标，让我们观察到感兴趣的动态变化。用时长仅为一分钟的视频展现两个星系碰撞的过程，可以压缩真实世界中持续数亿年之久的事件。同样，分子动态变化的时标也可以被延长为原来的数百万倍，以便研究人员理解原子和分子在几纳秒内发生的事件中的动态行为。

模拟也使得科学家和工程师可以观测他们感兴趣的任何量（比如，集成电路中晶体管的某个部分的行为），而在许多情况下，真实物理系统中的这些量都是无法观测的。我们显然不能测量恒星内部的温度，甚至不能测量地心的温度，但通过对这些系统进行模拟，我们可以确定太阳中心的温度大约是1 600万开氏度，地心的温度大约是6 000开氏度。足够精细的模拟器可以模拟我们感兴趣的任何变量，并监测它们的值。

有了可以调整时标、监测系统内部任意变量的模型，建模者就能从独特的视角了解系统的行为，并达到其他方法无法企及的深度。通过模拟形成蛋白质的原子团的行为，研究人员可以理解为什么这些蛋白质会执行某些功能及其背后的进化机制。这种深入理解有时比较容易实现，有时则很难（甚至不可能）实现，但通过模拟系统的行为，我们总能获得其他方法无法实现的重要见解。

模拟系统的能力之所以重要，第四个也是最后一个原因是，它能对尚未出现的系统行为进行预测。在建造一架新飞机之前，机械工程师可以模拟这架飞机的空气动力学行为，为设计决策提供参考，进而影响该飞机的最终设计。同样，计算机芯片的设计者也可以在芯片被

制造出来之前对其进行模拟。在这两种情况下，模拟都可以节省时间和金钱。

在模拟技术领域，生物系统模拟是一个新近的产物。直到不久前，我们对生物系统的一般性了解和对细胞的具体了解都还非常有限，以至于我们无法建立有效的模拟器。在进入20世纪的最后20年之前，我们对细胞的动态行为还不甚了解，完全不知道该如何构建切实有效的模型。对控制细胞行为的机制（从创造蛋白质的过程到多种细胞调控机制）的理解，为我们建立精细的细胞模型和细胞系统模型打开了一扇门。

单个细胞的动态行为非常复杂，利用现在的计算机还无法建立精细的原子水平的细胞模型。事实上，这个目标可能永远无法实现，从经济的角度看也不可行。但是，这并不意味着我们不能精确地模拟细胞和细胞系统。解决这个问题的关键是抽象和多层次分级模拟，这两个概念广泛应用于多种复杂系统，特别是超大规模集成电路的模拟。

抽象模拟的基本原理是，一旦很好地理解了系统的某个部分的行为，我们就可以在更高的层级上模拟它的行为，从而得到该部分的抽象模型。该模型又被用于更高层级的模拟，从而实现十分精确有效的多层次分级模拟。比如，我们常会建立单一晶体管的电平模型。晶体管由几百万甚至几十亿个原子组成，想要描述晶体管的电学性质，我们可能需要对晶体管的一小部分，在原子和电子水平上进行精细的模拟。对晶体管的这一小部分的模拟，通常是在非常精细的物理水平上进行的。在模拟时，半导体材料晶体结构中的原子和晶体杂质的原子，都需要纳入考虑。这些结构的行为，以及这些原子的电子云中的电子行为（受电场影响），都得到了十分精细的模拟，为成功描述这一小块晶体管的行为创造了条件。在深入了解并准确描述这一小块晶体管的行为之后，我们就可以为整个晶体管建立电平模型了，并利用该模型准确地模拟整个晶体管的行为，尽管它无法模拟单个原子或电子的行为。电平模型可以描述施加电压差后流经晶体管的电流量，也

可以描述晶体管的其他性质。这种模型还可以足够准确地模拟电路中的晶体管行为，而且它比建立该模型时进行的原子水平的模拟要快几百万倍。

利用这些晶体管模型，我们可以有效地模拟几千个甚至几百万个晶体管，虽然在电子和原子水平上进行全模拟目前还做不到。在实践中，我们通常不会对集成电路中的几百万（甚至几十亿）个晶体管进行精细的全模拟。更高级别的抽象模拟利用这些电平模型和已开发并投入应用的更加抽象的系统模型，描述包含数百甚至数千个晶体管的系统。例如，算术逻辑部件是计算机的核心，它包含几千个晶体管，但我们只需要描述它们得出一个计算结果所需的时间和电能（取决于操作类型和操作数类型）即可。因此，我们可以利用这种模型，模拟当该部件与存储器及其他部件连接起来时集成电路的行为。我们还可以利用存储器和其他部件的高层级模型，模拟几十亿个晶体管（每个晶体管都由几十亿个原子组成）的行为。考虑到电路中其他部分（比如电线和绝缘材料）也以一种不明显的方式被模拟，所以整个系统的模拟涉及10^{21}个原子。

那些可以有效模拟极其复杂的生物系统的技术也使用了这种方法，尽管到目前为止，它们的发展势头远不及集成电路模拟技术。研究人员正在努力开发生物系统的分子间相互作用（特别是DNA、蛋白质和代谢物之间的相互作用）模型，以应用于更高层级的抽象模拟，其中包括前文讨论的各种类型的生物网络。例如，基因调控网络可用于推导当细胞对特定化合物的攻击做出反应时，特定蛋白质浓度的动态变化。基因调控网络代表了一种高层级的抽象模型，它只关注细胞内部的蛋白质和其他分子的浓度。但是，这种模型可用于模拟当特定转录因子附着到基因编码区附近的DNA上时发生的现象。模拟一个（或一组）转录因子与一个DNA片断之间的连接，是一项非常复杂的任务，可能需要广泛地模拟与连接有关的分子动态变化。事实上，这种机制非常复杂，人们对它还不甚了解，所以现有的模型都不完整。但是，随

着研究的不断深入，可以精细地表现这一过程的模型必将出现。到那时，我们就可以利用这些精细的模型去建立高层级模型，展现转录因子的浓度与基因的转录水平之间的关系；我们还可以利用基因调控网络模型中的更高层级的模型，展现与这个过程有关的所有分子浓度的动态变化，从而对一个涉及数万亿原子的系统进行精细和有效的模拟。

通过将模拟生物系统不同方面的各种网络相互连接起来，以及使用越来越抽象的模型，最终我们有可能为大型细胞系统、完整的器官甚至是整个生物建模。一般来说，人们感兴趣的并不是建模并模拟生物体内的所有过程。在大多数情况下，研究人员关注的是通过建模和理解特定的过程或途径，弄清楚如何应对特定的功能障碍或疾病。人们感兴趣的许多需要模拟复杂生物系统的问题，都与疾病及其他衰弱情况有关。事实上，想要通过建模搞清楚每一种复杂疾病的发病机制和找到可能的疗法，需要成千上万的研究人员付出艰辛的努力。很多治疗方法都必须设计出特定的分子，去干扰致病的生物网络的行为。诸多癌症研究都把控制细胞分裂、细胞增殖的蛋白质和基因调控网络作为目标，探索如何阻止失控的细胞分裂行为。要理解逐渐摧毁或损害神经元细胞的神经退行性疾病，不仅需要建立精确的神经元网络模型，还需要建立模型来模拟控制每个神经元行为的分子机制。衰老本身并不是一种疾病，但如果我们构建出能更好地模拟细胞分裂机制和神经元网络行为的模型，就可以延迟（甚至有可能规避）衰老的影响。

因此，生物的组件模型将被用来理解和治疗许多影响人类的疾病和状况。一般来说，这些模型的使用不会引发复杂的伦理问题，因为它们都是生物组件的计算机模型，而且是为了治病救人。在大多数情况下，至少在研究的某些阶段，它们还可以减少新分子影响的相关研究需要做的动物实验的数量。但是，模拟整个生物的能力会带来一些引人关注的伦理问题，因为如果这些模型足够准确和精细，可以再现真实生物的行为，实际上就是一种仿真。迄今为止，人类在模拟复杂

生物方面取得的最显著进步或许是“开放蠕虫”项目，它的目标是模拟完整的秀丽隐杆线虫。

数字动物

把体长1毫米长的秀丽隐杆线虫作为研究简单多细胞生物的模型，在科学领域有着悠久的历史。如前文所述，它是最早进行基因组测序的动物。但在此之前的1963年，悉尼·布伦纳就提出将其用作研究动物神经发育的模式生物的建议。正是抱持着这个想法，20世纪七八十年代，布伦纳一直在英国剑桥大学MRC分子生物学实验室做相关研究。经过12年的努力，布伦纳完成了秀丽隐杆线虫的完整大脑结构的反向工程，并绘制出这个简单大脑中的所有神经元连接图。由于这种蠕虫的大脑结构完全是由基因组编码的，共享特定基因组的所有个体的大脑几乎完全一样，这让布伦纳的研究工作变得不那么复杂。

完成蠕虫大脑的反向工程需要付出艰苦的努力：先将若干蠕虫大脑切成极小的薄片，再用电子显微镜为它们拍摄大约8 000张照片，然后将每个切片中的每个神经元片段与相邻切片中的相应神经元片段连接起来（主要靠手工完成）。一篇长达340页的论文《秀丽隐杆线虫的神经系统结构》，详细描述了秀丽隐杆线虫大脑中的302个神经元和大约7 000个突触的完整连接图。论文的副标题更加言简意赅：蠕虫的头脑。理论上，我们可以根据这幅详细的连接图，创建一个非常精细而且行为与真实蠕虫十分相似的模型。事实上，开放蠕虫项目的目标就是构建一个完整的模型，不仅包含302个神经元和95个肌肉细胞，还包含蠕虫体内的其余1 000个细胞（更确切的数字是，959个体细胞加上雌雄同体蠕虫体内的大约2 000个生殖细胞，以及雄性蠕虫体内的1 031个细胞）。[最新电子书免费分享社群，群主V信 1107308023 添加备注电子书]

始于2011年的开放蠕虫项目必须克服巨大的困难，才能建立起有效的蠕虫模型。要确定蠕虫大脑的真实行为，仅靠布伦纳团队取得的大脑连接图远远不够。项目组需要获取更多的信息，比如，神经元（突触）之间的所有连接的精细模型，神经元的动态模型，以及影响大脑行为的其他变量的更多信息。但从理论上讲，仅靠细胞行为的进一步研究、先进的仪器技术和模型参数的微调，是不可能获得这些信息的。

开放蠕虫项目汇集了程序员（他们利用编程语言编写完整的模拟器），以及遗传学、神经元建模、流体动力学和化学扩散等多个领域的科学家。它的目标是建立所有肌肉和神经元的所有电活动的完整模拟器，建设可以模拟蠕虫体内的身体运动和物理作用力，以及蠕虫与其所处环境的相互作用的一体化模拟环境。

我们之所以要在这里讨论开放蠕虫项目，是因为该项目通过在计算机中仿真细胞过程，理论上可以实现完整模拟较复杂动物行为的目标。我们不妨暂时接受这样一个合理的假设：开放蠕虫项目将成功建立一个精细的秀丽隐杆线虫模型。如果在面对各种各样的刺激时该模型的行为与真实的秀丽隐杆线虫难以区分，我们就必须承认这个模拟蠕虫的程序通过了非常严格和具体的图灵测试；也就是说，这个计算机程序模拟的蠕虫，可以成功地骗过观测者。有人可能认为，蠕虫模拟永远不可能达到以假乱真的程度。但是，这个观点过于简单化了。相较于逼真的蠕虫模拟，利用标准的计算机图形技术在屏幕上实时呈现它的行为，并达到与真实蠕虫难以区分的程度，其难度显然微不足道。

在现实中，没有人会在意十分精细的蠕虫模拟是否与真实的蠕虫难以区分。但这种可能性给我们带来了一种更具挑战性的可能性：如果随着科技的发展，我们也能够同等精细地模拟其他复杂动物，会怎么样？如果我们可以精细地模拟老鼠、猫、猴子，或者至少是这些动物的重要部位，会怎么样？如果我们能模拟猴子，又有什么能阻止我

们模拟人类呢？毕竟，灵长类动物和人类的生物复杂性相当接近。最具挑战性的差异在于大脑，而它也是我们最想尝试模拟的器官。尽管人类的大脑有大约1 000亿个神经元，而黑猩猩的大脑只有70亿个神经元，但如果我们能实施反向工程并模拟黑猩猩的大脑，这样一个数量级的差异可能并不会成为我们的障碍。对此，持怀疑态度的读者也许会认为，这种可能性遥不可及，短期内不太可能发生。模拟有302个神经元和不超过1 000个细胞的蠕虫，与模拟更复杂的动物有着天壤之别，更不要说模拟有1 000亿个神经元和几万亿个细胞的人类了。

尽管如此，但我们必须记住这种可能性是存在的。模拟复杂动物是一个极具价值的目标，一旦实现，就会改变我们理解、操控和设计生物系统的能力。目前人类的能力仍然有限，但合成生物学的发展可能会改变这种状况。

合成生物学

随着我们对生物系统的深入理解，终有一天我们能够从零开始设计全新的生命形式。想要实现这个令人兴奋的可能性，我们就必须探索浩瀚的孟德尔图书馆，找到特定的DNA序列，然后利用细胞机制解释这些序列，最终创造出可存活的新生物。人们通常把合成生物学定义为设计、建造实用的生物装置和系统。最早使用这个术语的人可能是遗传学家瓦克劳·希巴尔斯基。1973年，在一次小组讨论中，当有人问希巴尔斯基“你在不大可能预见的未来想做什么”时，他回答道：

> 现在，我来谈谈“接下来要做什么”的问题。到目前为止，分子生物学研究还处于描述性阶段……但真正的考验要在我们进入研究的合成阶段之后才会到来。到那时，我们将设计新的控制构件，并将这些新模块添加到现有的基因组中，或者构建全新的基因组。这将是一个拥有无限扩展潜力的领域，我们可以几乎不受限制地构建出“新的更好的控制电路”或者……最终创建出其他“合成”生物体，比如“新的更好的老鼠”。……总的来说，我相信合成生物学领域的令人兴奋的新创意将层出不穷。

设计新的生物系统是一项复杂的工作。人造电子和机械系统由模块化部件以清晰的方式相互连接而成，每个模块具有不同的功能。而生物系统是复杂的奇妙装置，它其中的每个部分都以复杂和不可预测的方式与其他所有部分相互作用。经DNA编码的同种蛋白质，在大脑中执行一种功能，而在皮肤中则有可能执行另一种完全不同的功能。而且，同种蛋白质在不同的细胞中也可能会执行多种不同的功能。就我们目前掌握的知识而言，还无法从零开始构建一种全新的合成生物。

但是，我们掌握的技术可以创造出特定的DNA序列，并把它们插入自身DNA被清空的细胞。这些细胞机器将对插入的合成DNA做出解释，并为DNA编码的所有细胞创建新拷贝。现在，我们有可能根据给定的特化（specification），合成有数千个碱基对的DNA序列。随着技术的进步，我们还有可能以非常合理的成本合成更长的DNA序列。

由于我们已经充分理解了某些生物组件的功能，通过在设计好的DNA中插入相应的代码来复制这些组件，可以降低我们在孟德尔图书馆中寻找新生物的难度。最常用的DNA组件是汤姆·奈特发明的“生物积木”质粒。生物积木是为特定蛋白质编码的DNA序列，每种蛋白质在细胞中都有明确的功能。生物积木质粒被存储在一个标准生物组件登记库中，任何对设计新生物系统感兴趣的人都可以使用它们。事实上，在国际基因工程机器（iGEM）大赛等生物系统设计竞赛中，世界各地的成千上万的学生都会使用这个生物系统设计的“标准”。此外，还有一些本质上与生物积木类似的其他组件和方法，也得到了广泛应用。

人们已经通过合成生物DNA并将其引入工作的细胞机器，“重新创造”出某些现有生物。2003年，克雷格·温特研究所的一个团队合成了噬菌体Phi X 174的基因组，共包含5 386个碱基，它是第一种完成基因组测序的生物。这个团队利用合成基因组，组装出活的噬菌体。更重要的是，在2006年，这个团队构建出实验室支原体的合成基因组，并申请了专利。实验室支原体是衍生于生殖支原体的一种新型的最小细菌，他们这样做的目的是创建一种可以存活的合成生物。之所以选择生殖支原体，是因为它是已知的基因数量最少的生物。后来，研究团队决定使用另一种细菌，即丝状支原体，并成功地将这种细菌的合成基因组移植到另一种细菌的细胞（DNA已被移除）中。据报道，这种新细菌存活下来，并成功进行了繁殖。

诸如此类的努力至今还未创造出特征和行为完全不同于现有物种的全新物种，但毫无疑问，只要继续努力，最终必然会出现可以改变

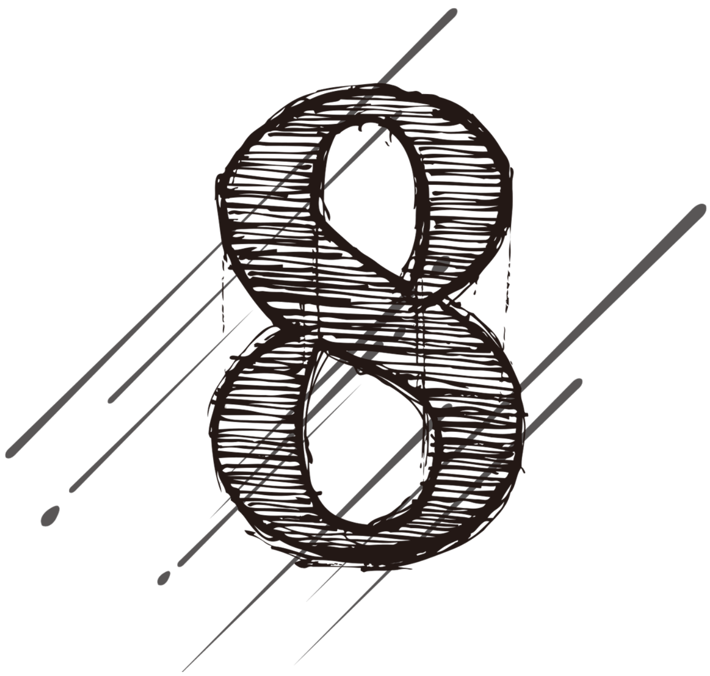
现有生命形式的新技术。改变生命形式有望解决多个领域的诸多问题。在理论上，新设计的细菌可以利用二氧化碳和阳光产生碳氢化合物，还可以清理石油泄漏，对抗全球变暖，以及执行其他许多任务。

如果说从零开始设计出可以存活的新型单细胞生物仍然需要很长时间，设计出复杂的多细胞真核生物就是更加遥远的目标了。创造不同的细胞类型和身体设计是一个非常复杂的发展过程，而且极其依赖DNA序列。只有更加全面地理解生物网络的运行方式，我们才有可能对现有的生命形式进行一些小“调整”，但我们无法从零开始设计出全新的物种。短时间内，独角兽和半人马都不会出现在我们的动物园里。在过去的许多年里，人类通过基因工程在现实世界中设计出马、狗等动物；在未来的许多年里，合成生物学将成为一项更加精妙复杂的基因工程。

在下一章，我们将就前几章描述的这些技术可能会如何改变世界，做出一些有根据的猜想和大胆的推测，但在此之前，我们需要更好地理解最复杂的器官——大脑——是如何工作的。事实上，创建精确的大脑模型，可能是我们在探索全面理解复杂生物行为的征途中面临的最大挑战。

第8章

大脑是如何工作的



自从克罗顿的阿尔克迈翁和希波克拉底认识到大脑是智能的源泉以来，人类就一直想知道大脑是如何工作的。尽管相关研究很早就开始了，但直到2 000多年以后，人们才普遍认识到智力和记忆都存在于大脑中。而在此之前的很长时间里，人们一直以为智能存在于心脏中；即使到了今天，我们仍然会说“用心”记住某些东西。

根据最新的估计，人体细胞的总数平均为37万亿个，脑细胞的总数约为860亿个，这意味着只有不到0.5%的细胞存在于我们的大脑中。但是，我们之所以是我们，这些细胞在其中起着至关重要的作用。它

们决定了我们的思维，如果没有它们，我们的记忆、个性和本质都将不复存在。

我们中的大多数人都对自己的身体或者特定的生理特征有着强烈的感情。但如果让我们做出选择，大多数人可能更愿意成为大脑捐献者，而不是被捐献者。当然，其前提条件是我们有能力进行大脑移植——未来几年里，我们可能会做这种尝试。

我们对大脑中的神经元和其他细胞结构的了解较晚。虽然显微镜大约是在16世纪末发明的，但由于神经元比许多其他类型的细胞更难观测，所以过了几百年显微镜才被成功地应用于脑细胞观察。

神经元非常小，从几微米到100微米不等。但就大小而言，它们和红细胞及精子差不多。神经元之所以很难观测，是因为它们致密地缠绕在一起，很难辨别单个细胞。承现峻在他2012年出版的开创性著作《连接组》中做了一个类比：大脑中的神经元——混成一团的细胞体、轴突、树突树和其他支持细胞——看上去就像一盘煮熟的意大利面。即使你用显微镜观察脑组织切片的表面，也不可能分辨出单个细胞，因为你能看到的只是一团纠缠不清的细胞成分。

1873年，卡米洛·高尔基取得了一项重大突破。他发现的脑组织染色法可以只对一小部分神经元染色，使它们从其他神经元中凸显出来。我们虽然不知道为什么高尔基染色剂只会使一小部分神经元可见，但它能帮助我们在显微镜下将神经元与它们周围的生物组织区分开。1904年，圣地亚哥·拉蒙－卡哈尔利用高尔基染色剂和显微镜，令人信服地证实了神经细胞是彼此独立的，他还发现了多种类型的神经元，绘制了大部分脑图。在发表研究成果时，拉蒙－卡哈尔一并附上了多幅神经元示意图（图8－1），其中一些沿用至今。

神经元及其分支的辨认，代表了我们在大脑理解方面取得的第一个重大进展。拉蒙－卡哈尔在他1901年出版的著作《我一生中的回忆》（*Recuerdos de mi vida*）中写道：“亲眼目睹铬－银反应的富

有启发性的神奇力量时，我感到十分惊讶，但科学界对这一发现却没有表现出任何兴奋之情。”

高尔基和拉蒙－卡哈尔因为他们的贡献共同获得了1906年的诺贝尔生理学或医学奖，但他们从未就神经细胞间的相互作用方式达成一致意见。高尔基认为神经元相互连接，形成一种超级细胞；拉蒙－卡哈尔则认为不同的神经元相互接触但彼此独立，通过某种未知的方式进行交流。20世纪20年代，奥托·勒维和亨利·戴尔证明神经递质参与了神经元之间的信号传递，最终确认拉蒙－卡哈尔的看法是正确的。但直到1954年，乔治·帕拉德、乔治·本内特和爱德华多·德罗伯蒂斯用刚发明的电子显微镜揭示了突触的结构，才提供了最终的证据。现在，我们知道突触分为化学突触和电突触，而且化学突触比电突触多得多。

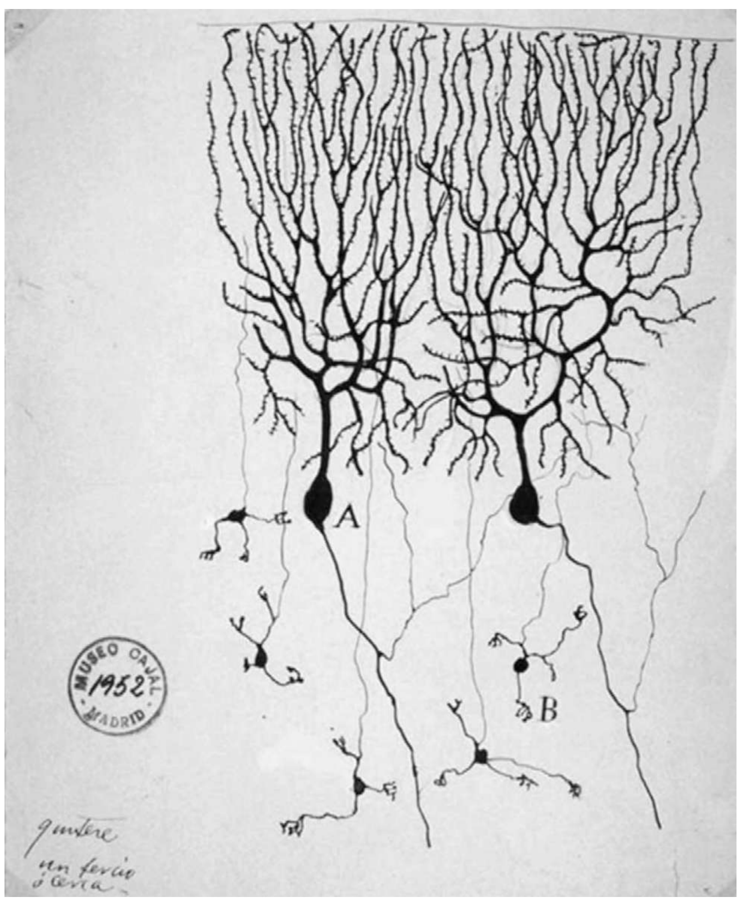


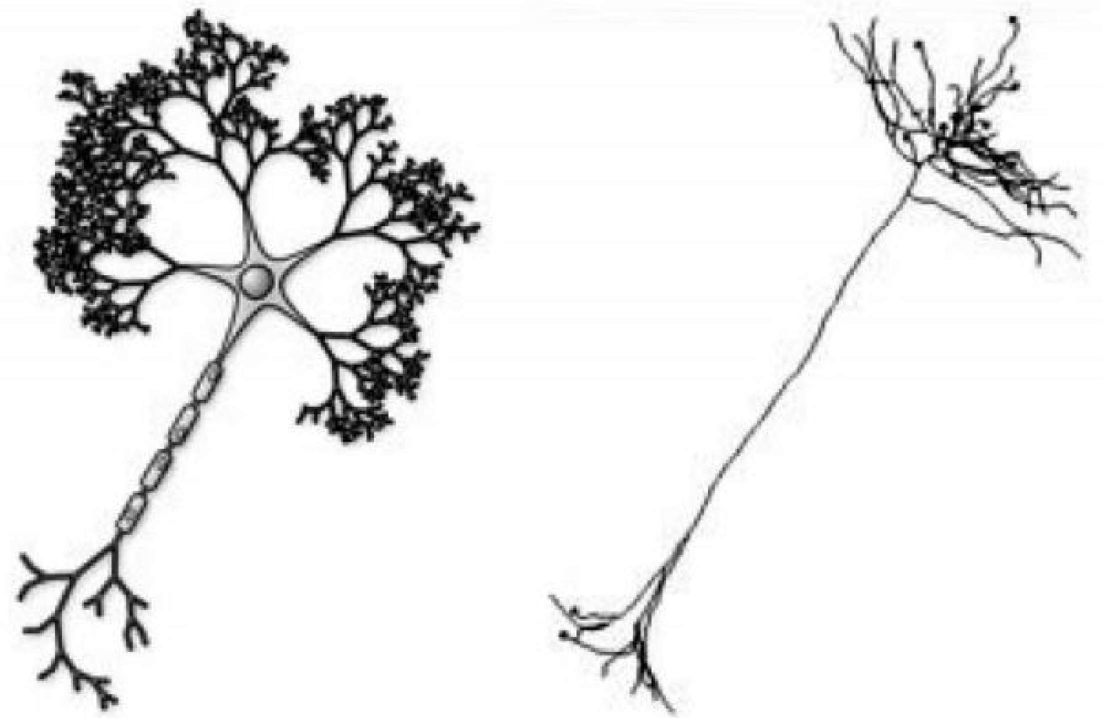
图8－1 圣地亚哥·拉蒙－卡哈尔绘制的鸽子小脑中的浦肯野细胞（A）和颗粒细胞（B）结构图

想要理解大脑是如何工作的，重要的是了解神经元是如何工作的，以及它们是如何构建出大脑这个极其复杂的系统的。

神经元是如何工作的

种类繁多的神经元可以根据形态、功能或位置进行分类。高尔基将神经元分成两大类：长有长轴突并用于长距离传递信号的神经元（I型）和长有短轴突的神经元（II型）。最简单的I型神经元（脊髓运动神经元就是一个很好的例子），由胞体和一条包裹着髓鞘的细长轴突组成。髓鞘有助于信号的传播，从细胞体上延伸出来的树突树负责接收来自其他神经元的信号。

普通神经元［图8－2（a）］有三个主要组成部分：树突、胞体和轴突。图8－2（b）描绘的是位于小鼠大脑主视觉区第四皮质的一个真实的神经元。神经元通常用树突树接收输入，然后以某种复杂的方式，将接收到的所有输入添加到胞体中；如果接收到足够的输入，神经元就会通过轴突发送输出。输出的特征是神经元放电，当神经元接收到足够的输入时便会发生。神经元放电是细胞膜产生的一种现象，细胞膜的电特性是从领鞭毛虫这种原始生物那里继承来的。



（a）　　　　　　　　　　（b）

图8－2 （a）尼古拉斯·鲁吉尔绘制的神经元示意图；（b）艾伦脑科学研究所重构的小鼠大脑主视觉区第四皮质的真实神经元图。树突末端的小圆圈表示构图终止的位置，而较大的圆圈表示胞体的位置

像其他细胞一样，神经元由包裹着脂膜的细胞机器以复杂的方式排布组成。在细胞膜内部，神经元拥有很多和其他细胞相同的细胞机器，包括不同类型的细胞器（比如线粒体）和一个细胞核，它们都在盐水溶液中游动。DNA复制和RNA合成就是在细胞核内进行的。但是，由于神经细胞的职能是以神经脉冲的形式将信息从一个地方传送到另一个地方，所以它们拥有与其他细胞不同的高度专门化的膜。神经细胞的膜是一个复杂的结构，含有很多帮助或阻止各种物质通过的蛋白质。就我们当下讨论的内容而言，值得关注的是相关的孔隙、离子通道和大小形状各异的离子泵。通道可以打开或关闭，因此它们控制着物质穿过细胞膜的速度。大多数离子通道只能渗透一种离子。离子通道允许离子沿着浓度梯度的方向移动，从高浓度区域移动到低浓度区域。离子泵是膜蛋白，它们积极地将离子泵入或泵出细胞，并利用细

胞能量（通常从ATP中获取），使离子以与浓度梯度相反的方向移动。一些离子通道，即所谓的电压依赖性离子通道，其泵送能力受膜两侧电压差（膜电位）的影响。

膜电位取决于神经细胞内外的带电离子浓度，通常为40~80毫伏，细胞外的膜电位为正。细胞浸泡在盐水溶液中，因为细胞膜两侧盐水的离子浓度不同，所以细胞内外的电位不同。通道中的离子沿某个方向流动，而其他离子则由离子泵泵送。在静止时，细胞膜两侧的电压差约为70毫伏。许多离子在细胞膜内外的浓度都不相同：钾离子（K^+）在细胞膜内部的浓度高于外部，而钠离子（Na^+）和氯离子（Cl^-）在细胞膜外部的浓度高于内部。这些不同浓度的带电离子会创造出膜电位。

由于神经元膜的存在，再加上离子通道和离子泵的这些特性，所以我们可以利用电路很好地模拟神经元的行为。有了这些一般性概念，我们现在就能考虑如何利用神经元的精细电路模型来模拟其行为。我们先尝试为一小片神经元膜建立电路模型。

为了进行信号传输，神经元采取的重要行为之一是，打开或关闭能让离子穿过细胞膜的离子通道和离子泵。通过为小片神经元膜建立精细的电路模型，并将这些模型相互连接起来，我们有可能获得整个神经元的有效模型。因此，小片神经元膜可以用一个简单的电路来模拟，如图8－3所示。细心的读者会注意到，这个电路和图3－2中用来说明麦克斯韦方程组的那个电路有很多相似之处。

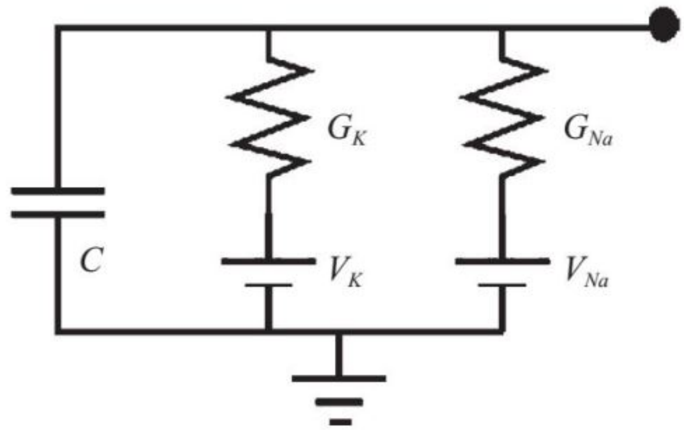


图8－3 小片神经元膜的简化模型电路图

在图8－3中，电压源V_K和V_{Na}代表离子浓度差异产生的平衡电位（本例只考虑钾离子和钠离子的情况）。电导G_K和G_{Na}表示可供离子通过的所有开放通道的综合效应。电容器模拟膜电容。在不影响模拟效果的前提下，我们可以假设神经元的外部电位为零。因此，神经元的被动膜可以用大量与之类似的电路来表示，每个电路对应一小片神经元膜，并通过信号传输方向上的电导相互连接。模拟钠离子和钾离子浓度的两个电压源，可以用电等效的电压源和电导（戴维南等效电路）来代替，最终的电路如图8－4所示。

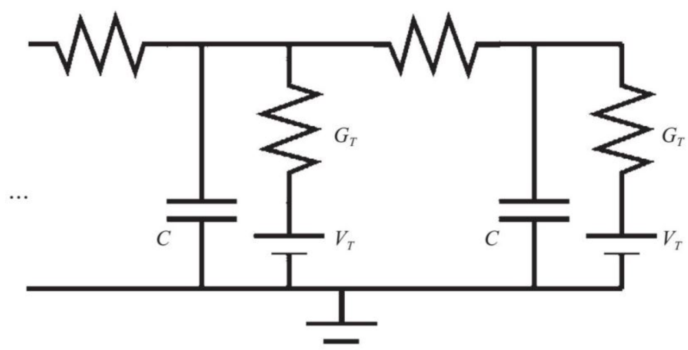


图8－4 由轴向电导连接的小片神经元膜的简化模型

艾伦·霍奇金和安德鲁·赫胥黎对神经元膜的电路模型进行了广泛研究。同真实的神经元一样，电路模型中也有多种不同类型的离子泵和离子通道。主动膜的行为比被动模型复杂得多，这是因为电导随时间和电压的变化而变化。但是，通过霍奇金、赫胥黎和许多其他研

究人员进行的广泛研究，我们对神经元膜工作原理的了解已经非常详细了。

在实验中，霍奇金和赫胥黎利用鱿鱼的巨大轴突，获得并确证了动作电位的离子机制模型。该轴突的直径长达一毫米，控制着鱿鱼喷水推进系统的一部分。它的尺寸使本来很难完成的实验和测量得以顺利进行。霍奇金－赫胥黎模型详细描述了电导根据特定方程随时间和电压变化的行为。后来，神经元的小片主动膜模型变成了图8－4所示的泛化模型，其中电导随时间和电压的变化而变化。每个电导都有一个特定的方程，电导的值是一个关于时间和电压的函数。利用该模型，霍奇金和赫胥黎于1952年成功地模拟了鱿鱼的巨型轴突中轴电位的产生过程，并准确地预测出模型中未明确包含的效应。特别是，他们利用从实验中获得的参数，对轴突传播信号的速度进行了较为准确的预测。许多其他研究人员还得出了特殊类型神经元的不同泛化模型。

图8－5展示的是人脑锥体神经元的小片主动膜的模型组件。图中所示的每个电导都对应于神经元膜中的一种特殊的离子电导机制。如果我们插入与这个电路有关的各个参数的具体时间和电压依赖性，并将多个这类电路（分别代表一小片细胞膜）彼此连接起来，就会得到一个完整神经元的电路模型。

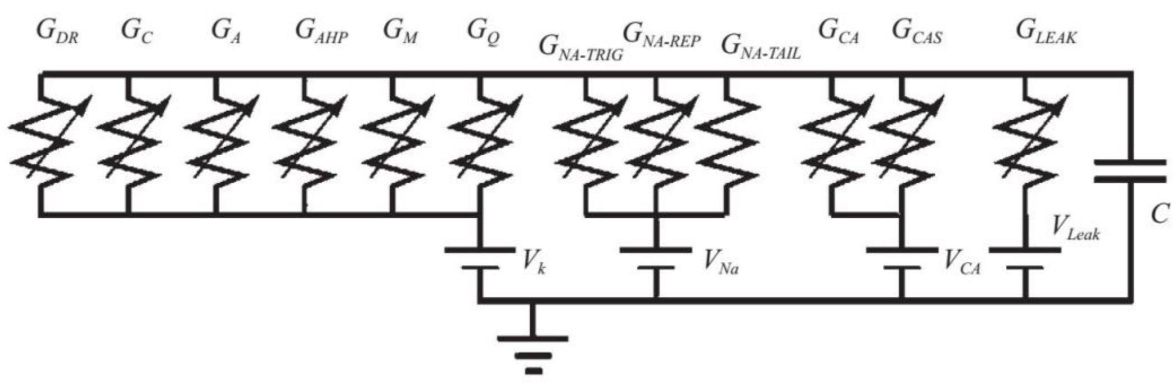


图8－5 小片锥体神经元膜的霍奇金－赫胥黎模型的完整电路图

当神经元不再接收其他神经元的输入时，膜电位的值是稳定的，神经元也不会放电。只在神经元通过连接神经元的突触——通常是突

触前神经元的轴突和突触后神经元的树突相连——接收来自其他神经元的输入时，放电才会发生。当有足够多的神经元通过轴突与该神经元的树突树相连并放电时，这些突触前神经元的电活动就会因为突触的活动而转变成电响应，进而导致接收神经元的膜电位上升（超极化）或下降（去极化）。

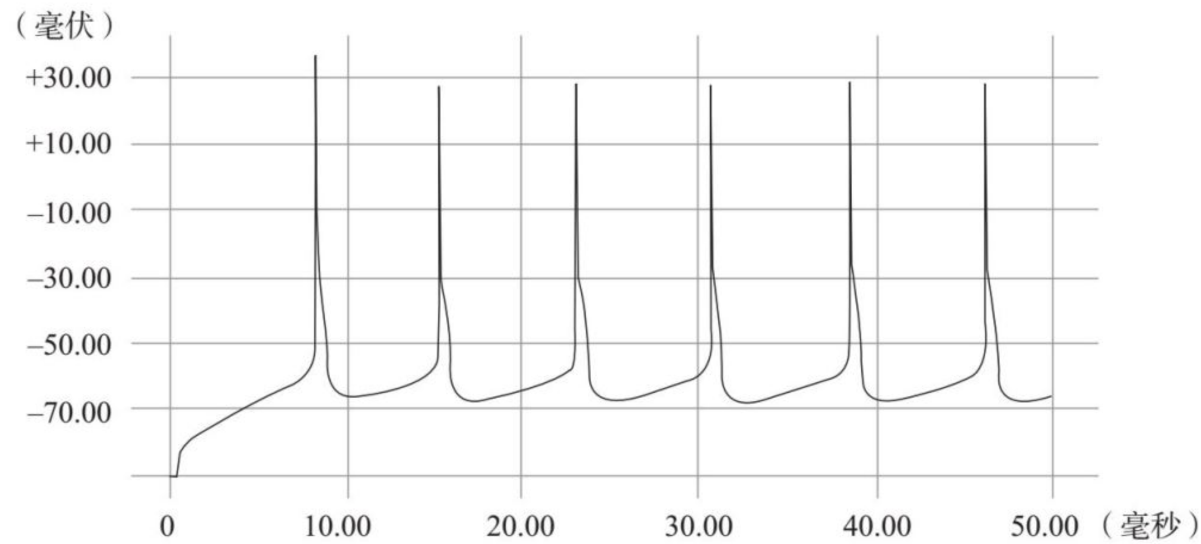


图8－6 锥体神经元的电特性模拟

如果接收神经元的一小片膜——通常位于轴突底部——被充分去极化，无论是在计算机模拟还是在真实的神经元中，我们都能观察到神经元膜的进一步去极化，这导致膜电位的快速自持下降（动作电位）。膜电位的这种快速自持下降对应于众所周知的神经元脉冲行为。当神经元产生脉冲时，膜电位突然去极化，在几毫秒内复位至接近静息状态的值。图8－6展示了神经元膜两侧产生的电压（运用模拟器得出的结果）。

当膜电压达到一定的电平时，它会经历一个再生过程，并产生一个脉冲。神经元在放电后有一个不应期，其间即使受激，它也无法产生其他脉冲。不应期的时长因神经元而异。图8－6所示的脉冲对应于小片神经元膜的电压随时间发生的变化。但是，由于这一小片膜与其相邻的小片膜彼此连接，所以它的去极化会导致相邻小片膜的去极化，进而使脉冲通过轴突传播，直到轴突末端的突触连接。主动膜的

特点是，一旦脉冲被引发，它的传播速度就只取决于轴突的几何结构。事实上，只要脉冲被引发，它的效应基本上就不受开始时它在胞体或树突树内部的波形影响。在某种程度上，它就是一个数字信号，要么存在，要么不存在。但是，神经元之间传输的信息是用脉冲的频率（或许还有脉冲发生的时机）来编码的。

当脉冲到达连接另一个神经元的化学突触时，它会迫使突触前神经元膜的突触小泡释放神经递质分子。这些分子附着在突触后神经元膜的受体上，并改变离子通道的状态。这些变化会引发离子流变化，使接收神经元的细胞膜去极化（兴奋性连接）或超极化（抑制性连接）。神经元膜两侧的两次电压下降的变化，分别被称为兴奋性突触后电位和抑制性突触后电位。化学突触是最常见的突触，位于突触前神经元膜和突触后神经元膜的间隙中，其大小为20~40纳米。

大脑还有电突触，电突触中有穿过目标神经元膜的通道，允许离子从该突触流向下一个神经元并传递信号。当突触前神经元的膜电位发生变化时，离子就可以在这些通道中移动。电突触传导神经冲动的速度比化学突触快，但它们不提供电增益。因此，突触后神经元的信号是起始神经元信号的衰减版。

凭借我们目前对神经元膜具体工作原理的了解，我们可以非常精确地模拟单个神经元或神经元网络的行为。这样的模拟要求我们详细了解每个神经元的结构和电性能，以及神经元是如何相互连接的，包括每个突触的具体特征。许多项目（比如，艾伦脑图谱）旨在为复杂大脑中的神经元建立精确的模型，这些项目的研究人员选择了一些特定神经元，用非常细的电探针将多种刺激物注入这些神经元，然后详细地测量它们的电反应。有了这些电反应，他们就可以创建出非常精确的神经元电路模型，并对它们进行调整。

与霍奇金和赫胥黎研究的神经元相比，复杂生物体的神经元要小得多，也更多样化。数十亿个神经元和数万亿个突触相互连接，形成了一个非常复杂的网络。了解这个复杂网络的详细结构（这可能是有

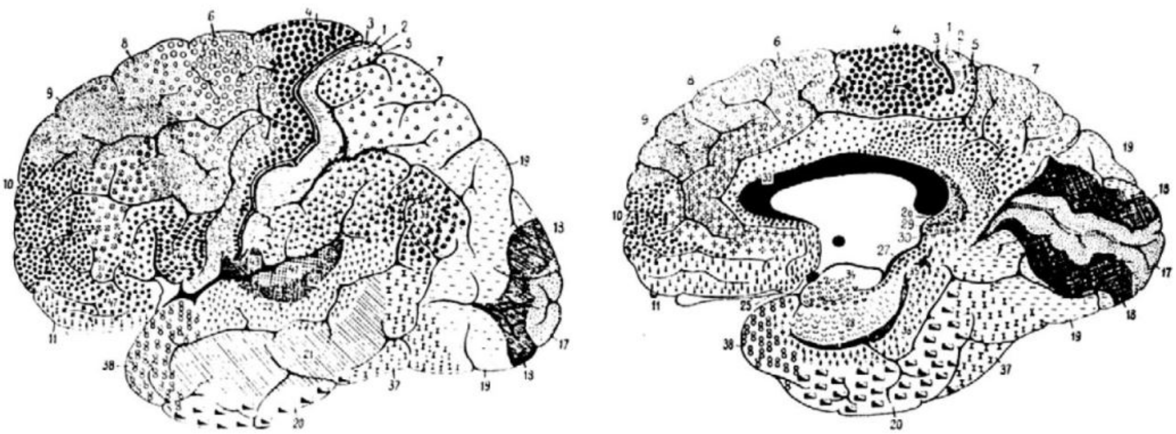
史以来最复杂的任务），可能会在医学、技术和社会领域引发翻天覆地的变化。目前，人们正在尝试用各种各样的方法去达成这个目标。

大脑的结构和组织

现代人脑有接近1 000亿个神经元，其中每一个都通过突触与其他多个神经元相连。据估计，人脑中的突触总数为10^{14}~10^{15}个，每个神经元平均拥有1 000~10 000个突触。但是，有些神经元拥有的突触数量远超10 000个。

神经解剖学使我们能够确定不同脑区的一般功能和特征。早在几百年前，人们就已经知道了灰质（主要由神经元的胞体组成）和白质（主要由神经元的轴突组成）的不同特征。脑的某些部位（包括皮层）主要由灰质组成，有些部位（包括胼胝体，即连接两个半脑的结构）则主要由白质组成。但是，脑的各个区域究竟是如何工作的，这在很大程度上仍然是个谜，尽管有些脑区比其他脑区更容易理解。人们通常认为脑可分为三个主要部分，即前脑、中脑和后脑，每个部分还可以细分成若干区域，每个区域都与身体行为的若干功能相关。

大脑（cerebrum，前脑的一部分）是人脑最大的部分，通常与较为高级的脑功能有关，包括记忆、解决问题、思考和感觉等。此外，它还负责控制运动。一般来说，与感官输入距离越近的脑区，就越容易理解。皮质是大脑最大的组成部分，具有特殊的意义，因为它参与较为高级的推理以及与认知和智能相关的活动。人们认为它比相对原始的脑部位更加灵活，也更容易调节。皮质是一层神经组织，厚度为2~4毫米，覆盖脑的绝大部分。皮质之所以形成褶皱，是为了在充分利用头骨内可用体积的前提下增加皮质的表面积。不同个体的褶皱模式相似，但也有很多细节上的不同。



（a）　　　　　　　　　　　　　　（b）

图8－7 布罗德曼区域示意图。（a）为侧面图，（b）为中轴面图
资料来源：史蒂芬·沃尔特·兰森，《神经系统解剖学》，1920。

解剖学家科比尼安·布罗德曼主要依据他在显微镜下观察到的组织细胞构成，对大脑皮质区域进行了定义和编号。在系统分析人类和其他几个物种的皮质微观特征的基础上，布罗德曼将皮质划分为52个区域（图8－7）。布罗德曼于1909年发表的研究结果，至今仍被用作皮质总图。

尽管许多其他研究提出了模式更详尽的图谱，但布罗德曼区域图仍然最广为人知，而且频繁被引用。100多年来，人们不断探讨、争论并改良布罗德曼分区图。布罗德曼根据神经元的组织结构定义了52个区域，其中多个区域后来被发现与各种皮质功能密切相关。例如，区域1~3为初级躯体感觉皮质，区域41、42与初级听皮质密切相关，区域17为初级视皮质。然而，有些布罗德曼区域只存在于非人灵长类动物中。在利用诸多技术（包括电极植入和各种成像方法）开展的脑研究中，“布罗德曼区域”这个术语得到了广泛应用。

基于成像技术的研究表明，脑在执行特定任务时，不同的脑区会变得活跃。依据这些研究结果绘制的详细脑图，可以帮助我们理解脑功能是如何分布的。一般来说，当脑执行若干不同的任务时，包括感觉处理、语言使用或肌肉控制，某个特定区域可能会变得活跃。

人们对视皮质区域的研究尤为充分。关于区域17的广泛研究，为我们了解大脑的工作原理提供了一个很好的示例。灵长类动物的大脑皮质初级视觉区（V1）与布罗德曼区域17重合，该区域执行视觉信息处理第一阶段的工作，与早期视觉处理、模式检测、轮廓感知、运动跟踪等功能有关。对该区域进行的广泛研究使我们充分理解了它的运行方式：它从视网膜处接收信息并进行处理，然后将这些信息转换为高级特征，比如边缘、轮廓和线性运动等，最后将它的输出传送到上游的视觉区域V2和V3。

通过对视网膜（其本身是一个极其复杂的系统）具体工作原理的广泛研究，人们充分了解了信号从视网膜到视觉皮质的流动过程。视网膜受体（视锥细胞和视杆细胞）检测入射光子并进行信号处理，其主要目的是检测圆心环绕特征（是中心暗周围明亮，还是正相反）。视网膜神经节细胞对这些圆心环绕特征敏感，它们通过视神经将神经脉冲发送到外侧膝状体核。外侧膝状体核是一个小的卵状脑部位，相当于一个中继中心：它对这些输入进行一些转换和信号处理，并获得三维信息；然后，它将处理后的信号发送到视皮质，也有可能发送到其他皮质区（图8－8）。

戴维·休伯尔和托斯坦·维厄瑟尔发明的用电极记录活体哺乳动物皮质活动的方法，为我们了解初级视皮质如何处理来自外侧膝状体核的信息创造了条件。研究表明，初级视皮质主要由对输入中的简单和复杂特征做出反应的细胞组成。

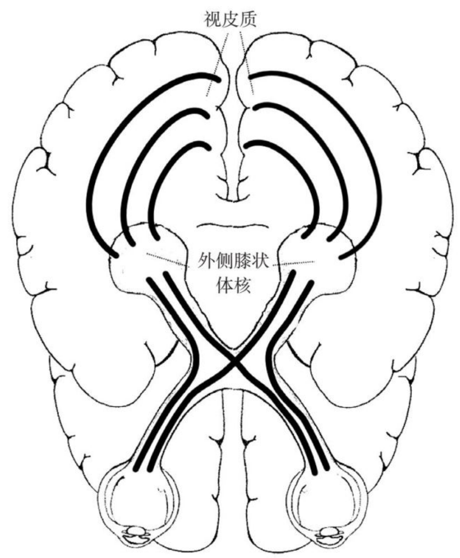


图8－8 脑图像处理的前几个阶段涉及的神经通路

休伯尔和维厄瑟尔的开创性研究——休伯尔于1988年出版的《眼睛、脑和视觉》一书对此进行了精彩描述——使我们进一步认识到，人类是通过从视网膜获取的图像中提取相关特征来感知世界的。随着信号越来越深入视觉系统，这些特征也变得越来越复杂。例如，眼优势柱（视皮质中的细胞群）对接收自一只眼睛的视觉刺激做出反应。眼优势柱是在视皮质区以条状排列的神经元群，优先对来自左眼或右眼的输入做出反应。这些柱状结构以条状模式排布在初级视皮质表面，跨越多个皮质层，负责探测不同的特征。在皮质表面检测到的特定特征不断变化，呈现出复杂的模式，被称为朝向柱。研究表明，简单细胞可以探测到视网膜的特定部分是否存在线条：要么是明亮区域中的暗线，要么是黑暗区域中的亮线。复杂细胞执行接下来的分析步骤。它们可以对扫过感受野的适当方向的线条做出反应，而不像简单细胞那样，只对位于其感受野的某一特定区域的静态线条敏感。人们发现，大脑皮质有许多不同功能的复杂细胞。当较长的边缘经过它们的感受野时，有些细胞的反应更加强烈，有些细胞对线条的末端敏感，有些细胞则对来自简单细胞的不同信号组合敏感。

从上述实验和其他实验中得出的初级视皮质结构，使我们对这一脑区的信号流动方式有了充分的理解。来自外侧膝状体核的传入神经

元主要进入皮质的第4层，将信息传递给具有中心环绕感受野的细胞和简单细胞。皮质的第2、3、5和6层主要由复杂细胞组成，负责接收来自第4层的各子层简单细胞的信号。

我们对其他脑区工作原理的理解程度都不及区域V1。尽管近几十年来研究人员对大脑进行了广泛研究，但到目前为止，我们对大部分脑区的作用及其运行方式的了解还相当笼统。进一步了解各脑区面临诸多障碍，比如，信号处理管道的上游区域非常复杂，解释大脑工作原理的方法缺乏，目前尚不能有效获取脑细胞行为的具体信息，等等。

但我们知道，我们不能将大脑视为一个简单地处理感官输入并将它们转化为行动的系统。即使在不处理输入的时候，大脑也会进行自发活动，产生我们通常所说的脑电波。这些波就算在静息状态下也会发生，根据它们产生的电磁信号的频率，可以分为α波（频率为8~13赫兹）、β波（13~35赫兹）、γ波（35~100赫兹）、θ波（3~8赫兹）和δ波（0.5~3赫兹）。这些频率范围是指示性的，并不具有唯一确定性。不同的脑电波与大脑的不同状态有关，比如深度睡眠、冥想和有意识的思考。尽管我们对自发神经活动及其产生的脑电波的作用进行了广泛研究，但仍然对这种活动起到的作用知之甚少。我们知道，即使在深度睡眠期间，大脑中的复杂神经活动也一直处于活跃状态，而且它们在认知过程中起着重要但大多未知的作用。这些活动是由神经元群的同步振荡产生的，神经元相互连接成大小不一的复杂反馈回路，有的是邻近神经元之间的连接，有的是在相隔很远的脑区之间形成的远程连接。

全面了解大脑的每个部位和每个神经元群的功能及行为，可能是许多年之后才有可能实现的目标。每个个体的大脑都不尽相同，对于不同的大脑，只有关于大脑的组织和功能的一般性规则才有可能相同。尽管我们还无法更好地了解不同脑区的详细功能，但努力了解大脑自我组织的一般机制，是脑科学当下的一个重要研究目标。

大脑的发育

说到对大脑的理解，我们必须记住一点：单靠一个人的聪明才智，可能永远无法全面理解某个特定大脑的具体工作原理。同样地，人类思维也不可能通过一点一滴的研究，去全面理解当今计算机的具体工作原理。但是，随着知识的进步，我们有可能对大脑的自我组织和发育的一般机制形成清晰的理解，就像我们对计算机采用的体系结构形成了大致了解一样。

理解大脑和理解计算机之间的这种相似性，对我们是有帮助的。虽然没人能记住现代计算机的具体行为，但计算机是由人类设计的，因此可以说人类了解计算机。人类对计算机的理解并不等同于我们知道任何一台或一组计算机中的每个晶体管的电压和电流。然而，人类了解计算机设计的一般原则，了解不同部件的相互连接方式，了解每个部件的行为，以及这些部件如何协同执行人类指派的任务。但对大脑来说，情况则复杂得多。尽管大脑的组织结构遵循某些一般性原则，但它们大多还未得到充分的理解。不过，我们掌握的知识足以让我们相信，大脑的组织结构是基因编码和大脑可塑性共同作用的结果。

显然，基因编码规定了细胞——由蛋白质、代谢物、水和其他成分构成——构建人脑的方式。人类基因组中的许多基因编码了人脑细胞的各种性质，并（以十分间接的方式）对人脑的结构进行了编码。但是，遗传信息并未对每个特定神经元和神经元之间的每个特定连接做出具体的规定。生物的基因组成控制着脑细胞的产生和复制，控制着大脑的整体结构，以及身体其他组成部分的结构。不过，在人类和

其他高等生物中，基因组成既不能控制神经元之间的特定连接，也不能控制大脑活动的模式。

对模式生物和人类的大脑发育的广泛研究，旨在进一步了解胚胎发育期间控制神经系统形成方式的细胞和分子机制。发育生物学领域的研究人员利用不同的模式生物，探索大脑的发育和自我组织方式。他们发现，通过化学梯度和其他机制，基因信息指引神经元将它们的轴突从一个脑区延展到另一个脑区。但是，神经元之间建立的具体连接模式过于复杂，不能由基因进行独特的编码，这些连接模式又过于动态，也不能由遗传因子进行静态的定义。基因组编码的是一个方法集，包括如何产生神经元，如何控制神经元细胞的分裂和繁殖，如何指导神经元轴突和树突树的延展，以及如何控制神经元与其他神经元建立连接。

当前研究的目的是了解发育过程的控制作用，即它如何控制神经元的产生和分化，神经元从诞生地到最终位置的迁移，轴突锥的生长和引导，轴突和突触后神经元之间的突触的形成，以及之后的许多分支的修剪。这些过程中有许多都是由生物的基因组成进行总体控制的，与特定神经元细胞的特定活动无关。

人脑的形成始于神经管。在妊娠期第3周，位于神经板上的神经前体细胞形成神经管。在妊娠期第8周结束时，多个脑区已经形成，不同的脑组织模式开始出现。

构成人脑的几十亿个神经元由神经前体细胞繁殖而来，神经前体细胞在繁殖时，通过分裂形成新细胞。神经元是成熟细胞，不会进一步分裂产生其他神经元。但是，神经前体细胞可以分裂很多次，每次分裂都会产生两个相同的神经前体细胞，新的细胞还能继续分裂。然后，这些细胞根据它们接收到的生化和遗传调节信号，产生神经元或胶质细胞。胶质细胞（也被称为神经胶质细胞）是非神经元细胞，其作用是维持大脑中的化学平衡，为神经元提供物理支持；它们还参与

髓鞘（一种包裹在轴突周围的绝缘体，可以提高轴突的信号传输速度，避免信号丢失）的生成。

人脑皮质形成新神经元的过程，大约在妊娠期的第16周完成，神经元生成后会分化成不同的类型。神经元主要是在后来成为脑室区的区域内产生的，之后会迁移到正在发育的新皮质和其他区域。神经元迁移的各种机制已经得到了广泛研究，但这一过程极其复杂，我们对它只是一知半解。

一旦神经元到达了它的目标区域，就会长出轴突和树突，以便与其他神经元建立连接。神经元在其附近区域建立致密的树突分支，并通过轴突的生长锥使轴突向外延伸。在化学梯度的引导下，生长锥朝着预定目标伸展。引导轴突生长的有些分子有吸引作用，有些有排斥作用，因此生长锥能够感知到一系列复杂的导向线索。一旦轴突到达它们的目标区域，就会与该区域的树突树建立突触连接。

在这一过程中发育的神经元以及它们建立的连接中，有很大一部分会在大脑发育的后几个阶段（出生前后）消失不见。相当一部分神经元会死亡，有些神经元即使没有死亡，它们建立的许多连接也会被剪除。在大脑的发育过程中，初始连接模式涉及的突触要比大脑在童年晚期达到更稳定状态后留存下来的突触多得多。总的来说，最终形成的突触总数可能比童年早期达到的峰值少一半。

神经元的迁移、生长锥的延伸和修剪等过程，在很大程度上受到遗传因子和生化因子的控制。不过，这些信息并不足以为完全成形的大脑中的各种连接模式编码。大脑的可塑性也起着非常重要的作用。在完全成形的大脑中，具体的连接模式在很大程度上取决于活动依赖性机制。在这种机制下，神经元的具体活动模式（源于感官体验和神经元自发激活）控制着新突触的形成和现有突触的修剪。

可塑性、学习和记忆

人脑是可塑的（也就是说可以改变），这种可塑性不仅体现在它的发育过程中，还贯穿人的一生。在新的经历发生时，可塑性赋予正常的大脑学习和修改其行为的能力。尽管可塑性在童年时期最强，但它是贯穿人的一生的一个基本而且重要的大脑性质。

每当我们看到或听到新事物时，每当我们形成新的记忆时，每当我们思考时，大脑的可塑性就会发挥作用。尽管我们还没有完全理解记忆的形成过程，但有显著的证据表明，长期记忆储存在大脑的连接模式中。短期记忆可能与大脑中的特定活动模式有关，但这些模式也与连接模式的变化（可能是短暂的变化）有关，这意味着在我们生命中的每一秒大脑的可塑性都是活跃的。

神经元之间的连接是动态的，它们的变化是突触可塑性作用的结果。突触可塑性不仅负责优化婴儿期早期新建立的神经回路，也负责在以后的人生中铭刻记忆，还负责形成几乎所有与学习及适应性有关的机制。对于控制突触可塑性（和大脑可塑性）的一般性原则，我们也只是一知半解。

学习机制不需要创建新的神经元，这个观点可能是拉蒙－卡哈尔最早提出的。1894年，他在英国皇家学会的克鲁年讲座上提出，记忆可能是通过改变现有神经元之间的连接强度形成的。

1949年，唐纳德·赫布延续了拉蒙－卡哈尔的观点，指出神经元可能会生长出新突触，或者发生代谢变化，从而增强它们交换信息的能力。他提出了两条简单的原理（后来，人们发现在很多情况下都会出现这两种现象）。第一条原理是，如果两个神经元被反复同步激

活，它们之间的联系就会加强。第二条原理是，如果两个神经元被反复依次激活，从第一个神经元到第二个神经元的连接就会增强。如果以一种相关的方式被激活，两个神经元之间的连接就会加强，我们现在称之为赫布法则。赫布在他1949年出版的《行为的组织》一书中首次提出了这个观点：

> 我们假设反射活动的持续或反复会导致细胞稳定性的持续提升……当细胞A的轴突与细胞B之间的距离很近，并持续或反复参与对B的激发时，其中一个或两个细胞的生长或代谢过程就会发生变化，使A激发B的有效性增强。

今天，这两条原理常被提及，以说明突触传递效率的变化是由突触前和突触后神经元的激发活动的相关性引发的，于是就有了“同时被激发的神经元连接在一起”这个众所周知的表述。事实上，这个表述比赫布最初提出的法则更具一般性，它意味着负责激发另一个神经元的神经元必须更活跃一点儿。这种基于相关性的学习理念，现在通常被称为赫布学习律，它可能在突触的可塑性中起着非常重要的作用。

休伯尔和维厄瑟尔通过用猫及其他哺乳动物做实验，研究了视皮质的自我组织。他们在这些动物的视皮质回路发育之前，剥夺了它们其中一只眼睛的视觉。在一只眼睛被剥夺了视力的猫身上，初级视皮质的柱状结构重新排列，占据了那只视力被剥夺的眼睛在正常情况下接收输入的区域。他们的研究结果表明，处理图像的皮质结构（比如简单细胞和复杂细胞）的发育取决于视觉输入。以更具体的视力剥夺方式进行的其他实验证实了这些发现。在其中一个实验中，猫从出生时开始，一只眼睛就只能看到水平线，另一只眼睛则只能看到垂直线。这些猫长大后，其视皮质中眼优势柱的排列呈现出相应的方式：视皮质细胞的感受野为水平走向或垂直走向（取决于对水平线和垂直线敏感的分别是哪只眼睛），并未发现对斜线敏感的细胞。因此，有

确凿的证据表明，在视皮质中发现的复杂排列方式在很大程度上可归因于活动依赖性的可塑性；在上述实验中，这种可塑性仅在婴儿期早期活跃。关于视皮质和专门负责其他感官的皮质区的许多其他研究结果，已经证实了休伯尔和维厄瑟尔的发现。不过，对于这种可塑性背后的机制我们仍然知之甚少，需要做进一步的研究。

人们认为有几种现象可用于支持突触的可塑性，其中最突出的就是长时程增强。人们对长时程增强进行了广泛研究，发现它与赫布法则有着密切的关系。长时程增强指的是突触强度长时间增加，以响应与突触前和突触后神经元有关的特定活动模式。与长时程增强相对的现象是长时程抑制。泰耶·洛默在兔子海马上发现的长时程增强，被视为支撑学习和记忆的细胞机制之一。

当受到充分的刺激时，许多脑组织就会发生长时程增强，但大多数相关研究都是在包括人类在内的多种哺乳动物的海马中进行的。当神经刺激的性质、强度和持续时间均恰当时，长时程增强就会表现为通路中神经反应的持续增强。研究表明，长时程增强的存在与记忆的产生具有相关性，阻断长时程增强的化学变化也会阻断记忆的产生。这一结果令人信服地证明，长时程增强即使不是最重要的长期记忆植入机制，至少也是有效机制之一。

另一种可能会对突触可塑性起到重要作用的现象是神经反向传播。尽管名称相似，但这种现象与第5章提到的反向传播算法之间没有明确的相关性。神经反向传播是指，（在大多数情况下）产生于轴突基部的动作电位，也会产生一个返回至树突分支的动作电位，尽管强度有所减弱。一些研究人员认为，就像多层感知器利用反向传播算法反向传播错误信号那样，这个简单的过程也可以发挥类似的作用；但到目前为止，还没有足够的证据证明已经发现了这样的机制。

突触可塑性并不是支撑大脑可塑性的唯一机制。虽然突触的产生、强化和破坏可能是形成大脑可塑性的最重要机制，但其他机制可能也参与其中并发挥作用。成体神经细胞不能繁殖，因此在成体中不

会产生新神经元（但是，海马内的一些干细胞可以通过分裂产生新神经元）。一般来说，由于没有新神经细胞产生，大脑可塑性不太可能通过最明显的机制——新神经元的产生——来实现。然而，现有的神经元发生迁移或长出新的突起（轴突或树突树）等机制已经被发现，它们可能对成年人的大脑可塑性起着重要作用。

对大脑可塑性和大脑发育的相关机制的进一步研究，最终将使我们对大脑如何进行自我组织形成更清晰的认知。只要理解了大脑发育背后的原理，我们就可以创建更好的模型，更有效地模拟基因控制大脑形成过程的非活动依赖性机制，以及赋予大脑可塑性的活动依赖性机制。

为了更好地掌握脑科学的现状，我们有必要关注该领域正在进行的项目，以及深入探索大脑内部奥秘所用的技术。

第9章

理解大脑

出于诸多原因，理解人类大脑的发育和可塑性现象背后的机制具有极其重要的意义。从医学的角度看，理解这些机制有助于我们找到方法去预防大脑退行性疾病，延长寿命，并增加享受高质量生活的时间。这些目标本身就足以证明，我们现在为了详细理解大脑如何运作而付出的巨大努力是值得的。不仅如此，从工程学的角度看，这些努力也是有必要的，因为对大脑运行机制的深入理解，将为我们在计算机中再现同样的行为提供不可或缺的工具。

为了增进对大脑的理解，我们需要全面、深入地了解大脑的内部组织结构，以及大脑的激活模式是如何编码想法的。仪器技术的进步为我们观察活体大脑提供了许多方法。我们的目标是，在不干扰大脑行为的前提下，提升我们对与大脑的创建和运行有关的结构和机制的认识。有些技术会损坏大脑，但可以提供关于神经元结构和连接的更精细的补充信息。

少数大型研究项目和大量小型研究项目，都以获得关于大脑结构和行为的信息为目标，这些信息不仅可以揭示大脑是如何工作的，还有助于我们理解其他重要的东西。

美国的人类连接组项目（HCP）耗资数百万美元，目的是收集大量的神经数据，并提供一个导览数据的用户图形界面。他们计划通过建立“网络地图”来获取关于活体人脑的新知识，“网络地图”旨在提供人类大脑内部的结构性和功能性连接的相关信息，并利用这些知识提升我们对神经疾病的理解。美国的另外一个重要项目——开放连接组项目（OCP），旨在创建公开的源于多种生物的连接组数据。

此外，美国还于2013年宣布启动“通过推进创新型神经技术开展大脑研究”（BRAIN）计划，也称人脑活动图计划。这项耗资数十亿美元的研究计划的目标是，绘制出人脑中每个神经元的活动。它的想法是，通过加速新技术的开发和应用，获取新的大脑动态视图，首次展示出单个细胞和复杂的神经回路是如何相互作用的。

欧洲的人类脑计划是一项十年计划，预算超过10亿欧元，主要由欧盟资助。它于2013年启动，由洛桑联邦理工学院负责相关协调工作。这所大学还曾负责协调蓝色大脑计划，该计划取得了迄今为止关于大脑模拟的最重要的研究成果。人类脑计划旨在通过整合健康大脑和患病大脑的相关数据，实现对人类大脑的统一的多层级的理解。它聚焦于识别大脑相关特征时必须获取、存储、组织和挖掘的数据，它的主要目标之一是，开发基于大脑电路和计算原理的新型神经形态技术和神经机器人技术。

日本的大脑研究计划（Brain/MINDS）是一个耗资数百万美元的项目，于2014年启动，旨在通过非人灵长类动物的大脑，更好地了解人类大脑的工作原理。和其他脑计划一样，日本的大脑研究计划的主要目标之一是，阐明与脑部疾病有关的机制。它的想法是，以狨猴（一种小型灵长类动物）作为模式生物，研究导致脑部疾病的认知和神经机制。（老鼠等更简单的模式生物，在进化程度上可能与人类相距甚远，无法为人类大脑研究提供合适的平台，而使用狨猴作为模式生物可能有助于摆脱这个限制。）

艾伦脑科学研究所是一家非营利性的私人研究机构，开展了若干旨在了解人类大脑的工作原理的项目。正在开发过程中的“艾伦脑图谱数据库”是一个高度综合性的信息系统，整合了通过活体脑成像、组织显微镜和DNA测序收集的数据，记录下小鼠、非人灵长类动物和人类大脑的许多不同信息。该系统可以提供某些基因活跃于哪个脑区、大脑连接数据，以及特定神经元细胞的形态和行为数据等信息。

脑成像技术

上述项目和研究类似问题的许多其他项目，都利用各种技术去推导人类、灵长类动物和其他模式生物大脑的详细信息。其中许多项目都会使用同样的技术去获取关于大脑结构的信息，在某些情况下，它们甚至会利用这些技术去获取关于神经元水平连接性的信息。

接下来，我们将考虑使用成像方法，以非侵入性方式获取工作大脑的相关信息，并观测它们的行为。这类方法叫作神经成像。当我们的目标是获取脑组织的三维结构和组成的详细信息时，我们可以把大脑看成是大量的体素。体素是二维图像中像素的三维等价物，更准确地说，体素是一个三维长方体，相当于体积的基本元素，其尺寸取决于成像技术。长方体的棱长短不一，从几纳米到一厘米甚至更长，这取决于技术和具体应用。平均而言，体素越小，包含的神经元就越少，对应的神经元活动水平也越低。因此，体素越小，利用成像技术观察电活动或化学活动来获取准确的神经元特征信息的难度就越大。体素越小，所花的扫描时间也越长，因为对许多技术而言，扫描时间随体素数量的增加而增加。根据现有的技术，用于活体大脑成像的体素通常包含几百万个神经元和几十亿个突触，实际数量取决于体素的大小和成像的大脑区域。体素通常按平面或切片的形式并列排布，以获得关于大脑的完整三维信息。许多用于大脑成像的技术也被用于身体其他部位的成像，尽管有些技术是专门针对脑组织的具体特征开发的。

神经成像依据不同的物理学原理，获取关于大脑结构和行为的详细信息。神经成像可分为两大类：结构成像（获取大脑结构的相关信息，包括导致大脑结构发生明显变化的疾病信息）和功能成像（获取

大脑功能的相关信息，包括用于诊断影响大脑功能但可能不会影响大脑结构的疾病信息）。

已经应用于神经成像的技术包括：计算机断层成像（CT），近红外光谱（NIRS），正电子发射断层成像（PET），不同类型的磁共振成像（MRI），脑电描记术（EEG），脑磁描记术（MEG），事件相关光信号（EROS），等等。

1971年，伦敦阿特金森－莫利医院的戈弗雷·豪斯菲尔德首次使用了计算机断层成像，这种成像技术利用计算机处理X射线图像，将虚拟的组织切片相互叠加，产生三维体层图像。频率在3×10^{16}赫兹到3×10^{19}赫兹之间的X射线，被广泛应用于物体的内部成像，因为它们可以深入动物组织，但穿过不同材料后会有不同程度的衰减。

体层成像使研究人员无须切开大脑就能看到大脑（和其他组织）的内部情况。计算机算法对接收到的X射线图像进行处理，并利用放置在身体外部的传感器获得的一系列二维图像生成器官内部的三维图像。这些图像通常是围绕着一个旋转轴拍摄的，因此被命名为计算机轴向断层成像。计算机断层成像数据可由计算机进行处理，以突出显示X射线束经过不同身体组织后的不同衰减程度。计算机断层扫描可以使用造影剂，也可以不用（造影剂注入生物体内后，会使特定的器官或组织在X射线下看得更清楚）。在计算机体层血管成像中使用造影剂，可以很形象地显示大脑中的血管结构。

虽然X射线照片的分辨率与标准照片相当，但计算机断层成像只能达到几分之一毫米的空间分辨率。计算机断层扫描只需要几秒钟，却会使受试者暴露在潜在的损伤性电离辐射之下。因此，虽然计算机断层扫描可以提供关于大脑宏观结构的重要信息，但在研究大脑行为时并不常用。计算机断层扫描主要用于确定大脑结构发生的与活动水平无关的变化。

因为在神经元放电时组织的特征会发生轻微的变化，所以利用成像技术可使大脑的活动可视化。最显著的效应之一是血流量的变化。当某个脑区被激活时，该区域的血流量会很快发生变化。有些成像技术（包括近红外光谱、正电子发射断层成像和磁共振成像）利用血流量的变化来探测大脑活动。依据水、氧合血红蛋白、脱氧血红蛋白吸收可见光和近红外光的事实，我们也可以获取神经元活动的位置信息。

血氧浓度会随着神经元的活动水平而变化，要使神经元在被激活并放电后恢复到最初的极化状态，就需要穿过神经元细胞膜来回泵送离子，这个过程需要消耗化学能量。激活离子泵所需的能量主要来自血液携带的葡萄糖。血流量增大，才能运输更多的葡萄糖，并以红细胞中的氧合血红蛋白分子的形式带来更多的氧。血流量增加发生在活跃神经元周围2~3毫米的地方。通常情况下，血液带来的氧超过葡萄糖燃烧时消耗的氧，造成该脑区脱氧血红蛋白的净减少。尽管我们可能预期血氧浓度会随着神经元被激活而下降，但其动态变化比我们想象的更复杂。在神经活性增加后，血氧浓度确实会立即下降，但在随后一段时间内血流量增加，而且补充量超过增加的需求量，因此在神经元被激活之后，血氧浓度实际上会增加。这一现象被称为血氧水平依赖效应，最早报告这个现象的人是小川诚二。它能改变放电神经元附近血液的性质，因为它可以提供大脑不同区域的血流量信息，所以可用于探测和监测大脑活动。血氧水平依赖信号的量值在几秒钟后就能达到峰值，然后回落到基准水平。有证据表明，血氧水平依赖信号与该区域神经元输入活动的关系比与输出活动的关系更密切。在皮质的某些部位，轴突较短，它们的末端在神经元主体附近。由于体素并未小到足以区分神经元不同部位的程度，血氧水平依赖信号是与神经元的输入相关还是与输出相关，没有多大影响。在其他脑区，轴突较长，因此输入和输出活动之间的差异可能非常显著。

由于血氧水平依赖信号弱，并且容易受到多种噪声的干扰，所以我们很难准确测量它的水平。在实践中，我们需要用复杂巧妙的统计

程序来恢复基础信号。用颜色对整个大脑或特定脑区的活动水平进行编码，我们就可以形象地观察到大脑的激活情况。通过监测血氧水平依赖信号，我们可以将活跃部位锁定到几毫米以内的范围，时间分辨率为几秒。通过使用除血氧水平依赖信号以外的其他生物标志来提高空间分辨率和时间分辨率的替代技术正在开发中，但它们也会受到其他限制。因此，今天使用的绝大多数技术都使用血氧水平依赖信号来代表大脑活性。

近红外光谱是一种基于标准电磁辐射的应用技术，它使用的光谱部分与计算机断层成像技术不同，频率范围为1×10^{14}赫兹到4×10^{14}赫兹。近红外光谱可用于研究大脑，是因为身体组织对近红外光子的传输和吸收，能反映出血红蛋白浓度的变化。近红外光谱频率范围内的光在被漫射和探测到之前能射入组织内几厘米，因此通过测量血氧水平依赖信号，我们可以利用近红外光谱不侵入地监测大脑功能。做近红外光谱测量时，需要先将适当频率的光子发送到人脑中，并探测漫射光，然后利用计算机算法计算引起光子漫射的物质密度。

光源发出的光进入组织，再从组织中漫射出来，并被探测器接收，因此近红外光谱技术可以敏感地探测到光源和探测器之间的组织体积。由于近红外光在被漫射之前只能穿透人脑几厘米，所以通常将光源和探测器放置在头皮上，两者相隔几厘米。它产生的信号主要被用于完成皮质表层的成像。近红外光谱是一种非侵入性技术，可以测量血流动力学信号，时间分辨率为100赫兹或更高，不过它无法规避血氧水平依赖信号反应缓慢的缺点。功能近红外成像在成本和便携性方面优于磁共振成像及其他技术，但它不能测量颅骨内几厘米以下的皮质活动，而且空间分辨率很低。近红外光谱在人类皮质的功能制图方面的应用也被称为扩散光学断层成像。

正电子发射断层成像技术会将不稳定同位素发射的粒子注入血液，它是基于宾夕法尼亚大学的戴维·库尔、卢克·查普曼和罗伊·爱德华兹在20世纪50年代的研究成果开发的。1953年，戈登·布朗内

尔、查尔斯·伯纳姆和他们在麻省总医院的团队，演示了正电子发射断层成像技术的首次医学应用。

使用正电子发射断层成像时，需要将含有放射性核的生物活性分子注射到人体内，然后探测该放射性核释放的正电子在衰变过程中发射出的高能光子对。正电子是一种亚原子粒子，与电子质量相同，但带有正电荷。事实上，它是一种反物质，即反电子。当普通物质中的放射性核素发生衰变时，会释放出正电子。正电子移动很短的一段距离后，会遇到电子。正电子与电子发生湮灭后会产生两个高能光子（能量为511千电子伏），它们从湮灭的位置离开，并沿几乎同一条直线上的相反方向运动。随后，经过计算机的自动分析，就可以构建出最初注射到体内的放射性核素（示踪剂）浓度的三维图像。

正电子发射断层成像最常使用的生物活性分子是葡萄糖的类似物——氟代脱氧葡萄糖。它可以反映成像示踪剂浓度与组织代谢活性之间的相关性，组织的代谢活动又会导致葡萄糖摄取量增加，并产生血氧水平依赖信号。正电子发射断层成像最常见的应用是检测癌组织，它与氟代脱氧葡萄糖的配合使用效果很好，因为癌组织的结构与非癌组织不同，可以在正电子发射断层显像中产生明显的特征标记。

当氟代脱氧葡萄糖用于成像时，每个氟代脱氧葡萄糖分子中的正常氟原子被放射性同位素氟－18的一个原子替代。在衰变（β+衰变）过程中，细胞核中的质子被中子替代，并释放出一个正电子和一个电子中微子（图9－1）。同位素^{18}F的半衰期是110分钟，其间，氟－18原子中有一半会释放出一个正电子并衰变成稳定的氧－18原子。

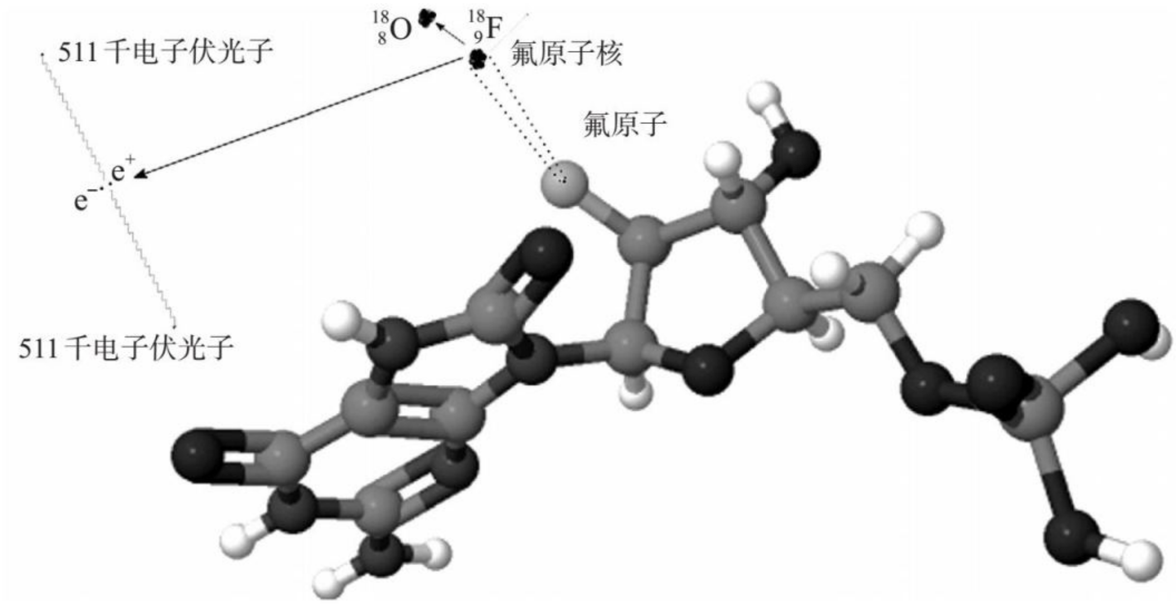


图9－1 氟－18原子核产生两个高能光子的衰变过程示意图

有一种神经成像方法是，将被标记的氟代脱氧葡萄糖注射到血流中，等待半个小时左右，以便未被神经元利用的氟代脱氧葡萄糖离开大脑。留存在脑组织中的被标记的氟代脱氧葡萄糖因为代谢作用而被困在脑组织中，其浓度可以表明那段时间内大脑中的哪些区域处于活跃状态。因此，根据可观测的正电子衰变数量测得的氟代脱氧葡萄糖在脑组织中的分布情况，是反映葡萄糖代谢水平的一个良好指标，而葡萄糖代谢水平又可以反映神经元的活性。不过，如果大脑活性与持续时间仅为几秒钟的心理变化相关，那么出于时标的考虑，该方法并不适用于相关研究。

但是，利用血氧水平依赖效应，正电子发射断层成像可以获取时间分辨率更高的大脑活性的相关证据。人们还利用半衰期比氟代脱氧葡萄糖短的其他同位素，比如碳－11（半衰期为20分钟）、氮－13（半衰期为10分钟）、氧－15（半衰期为122秒钟），合成各种各样的活性分子。想要使用寿命短的同位素，就需要在附近放置一台回旋加速器来产生不稳定的同位素，以便在大部分同位素发生衰变之前投入使用。

尽管上述技术都很重要，但现在研究宏观脑行为的最常用技术是磁共振成像。在雷蒙德·达马迪安和赫尔曼·卡尔的研究成果的基础上，纽约大学石溪分校的保罗·劳特布尔提出了磁共振成像理论。

磁共振成像技术建立在当氢核暴露在电场和磁场中时发生的一种物理现象的基础之上。氢核其实就是质子，它们具有一种叫作核自旋的内禀属性，它们可以像小磁体一样，顺着外加磁场的方向排列。施加磁场后，一小部分氢原子核（大多存在于水分子中）的方向就会与外加磁场的方向一致。在经历一个特定频率（谐振频率，其大小取决于磁场强度）的衰减振荡后，氢核的排列方向就会向磁场方向靠拢。强度为1特斯拉的磁场，其谐振频率为42兆赫，并且随磁场强度呈线性增加。现有的人体成像设备使用非常强（1.5~7特斯拉）的磁场，相比之下，地球表面的磁场强度为25~65微特斯拉，是人体成像设备所用磁场的十万分之一。

如果以适当的频率连续不断地轻推指南针，指南针的指针就会偏离南北方向。同样地，如果给这些氢核施加一个频率与氢核的谐振频率非常接近的无线电波，这些氢核的排列方向就会偏离外加磁场的方向。当这个激发过程终止后，氢核的排列方向再次与磁场方向一致，并以它们特定的谐振频率振荡。在此过程中，计算机可以探测和处理它们发射的无线电波，并用其创建三维图像。在这个图像中，不同类型的组织可以通过它们的结构和含水量加以区分。

实际上，由于原子发射的无线电波的波长太长，所以当原子以42兆赫或某个同样数量级的频率振荡时，我们不可能精确探测到它的位置。出于物理限制的原因，我们在探测无线电波的波源位置时，会存在一种无线电波波长量级的不确定性（我们在这里考虑的无线电波波长为几米）。不过，通过调节磁场，让它的强度随时间和不同的空间维度而改变，就可以精细地控制原子的振荡频率，使其依赖于原子所处的特定位置。这种调节使每个原子核发射波的频率随其位置和时间

的变化而变化，这样一来，它的位置就会被探测器探测到。图9－2展示了如何通过控制磁场来确定振荡氢核的位置。

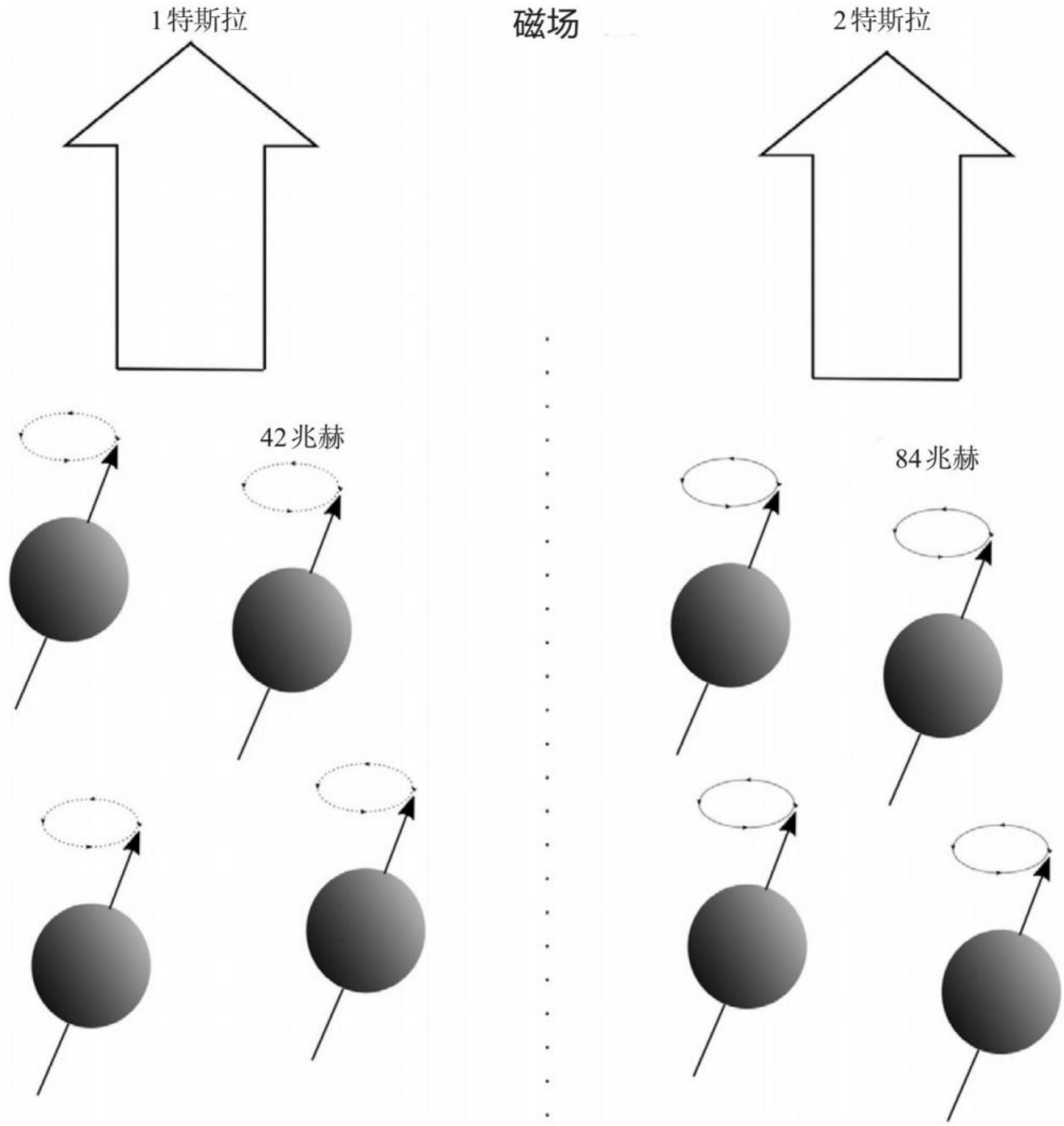


图9－2 利用质子（振荡频率取决于磁场强度）确定氢原子的精确位置示意图

磁共振成像在空间分辨率和时间分辨率方面存在一定的局限性。因为需要获得足够强的信号，所以体素不能太小。磁场越强，体素就越小，但即使是7特斯拉的磁共振成像仪也无法获得体素远小于1立方

毫米的高质量图像。不过，随着技术的发展，我们可以预期这种分辨率会提高，从而使磁共振成像能够获得体素小得多的图像。

根据不同的协议，我们可以利用磁共振成像的物理学原理获取不同的信息。磁共振成像最简单和最常用的形式是获取脑静态图像，帮助诊断导致大脑宏观结构发生变化的肿瘤或其他疾病。由于不同组织的磁共振成像标记不同，所以这些变化清晰可见。图9－3显示的是三个大脑切片的图像。图（a）是靠近头骨顶部的一个切片，皮质褶皱清晰可见。图（b）显示的是另外一些皮质褶皱，也可以看到侧脑室（产生脑脊液的部位）和胼胝体（大脑中最大的白质结构）。图（c）显示的是一个明显异常的结构（良性脑瘤），位于图像右侧靠近中心的位置。

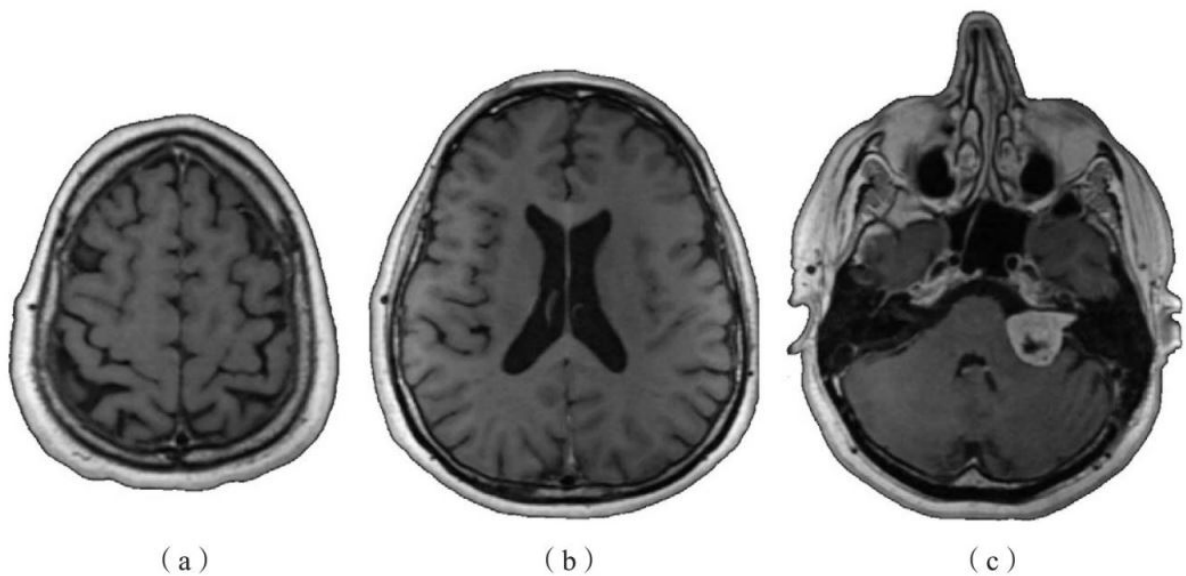


图9－3 通过磁共振成像技术获得的脑切片图像

利用磁共振成像技术，我们还可以获得关于大脑行为和结构的其他信息。功能磁共振成像和扩散磁共振成像是两种重要的磁共振成像技术。

近几十年来，功能磁共振成像在脑研究领域得到了广泛应用。1991年，麻省总医院艾瑟尼拉－马蒂诺中心的杰克·贝利维尔率先提出了功能磁共振成像技术。贝利维尔利用注射到血流中的造影剂，获

得一种与特定区域的大脑活动水平存在相关性的信号。1992年，邝健民、小川诚二和彼得·班德蒂尼宣称，血氧水平依赖信号在功能磁共振成像技术中可以直接用来表示大脑活性。

富氧血红蛋白和贫氧血红蛋白的磁性不同，导致氧合血和脱氧血的磁共振信号不同，因此血氧水平依赖效应可用于功能磁共振成像。在血液含氧量较高的地方，磁共振较强；在血液含氧量较低的地方，磁共振较弱。血氧水平依赖效应的时标在很大程度上可以决定时间分辨率。对于功能磁共振成像，血流动力学反应的表现是先迅速上升，在4~6秒后达到峰值，然后迅速下降，整个过程持续超过10秒。功能磁共振成像探测到的信号比触发信号的神经元事件慢1~2秒，因为血管系统需要花点儿时间才能响应神经元对葡萄糖的需求。

功能磁共振成像的时间分辨率可以满足多种大脑过程研究的需要。神经元活动的时间从100毫秒到几秒钟不等，较为高级的推理活动，比如阅读、交谈，可能要花几秒钟到几分钟的时间。因为功能磁共振成像的时间分辨率为几秒钟，所以这种技术可用于持续时间从几秒钟到几分钟的大脑过程实验研究，但由于需要重复做实验以提高信噪比，所以实验可能会持续将近一小时到几个小时不等。

扩散磁共振成像是产生生物组织结构的磁共振图像的另一种方法，图像表现的是分子（通常是水分子）扩散的局部特征。扩散磁共振成像技术最早是由丹尼斯·雷比昂于1985年提出的，它基于这样一个事实：我们可以制造出对分子运动敏感的磁共振成像仪，去展现与组织结构相关的微观差异。扩散加权成像和扩散张量成像是扩散磁共振成像采用的两种特定技术。

扩散加权成像得到的图像强度，与组织体素中水分子的随机性布朗运动有关。虽然组织结构与扩散之间的关系非常复杂，但细胞组织的密度越大，扩散系数就越低，因此扩散加权成像可用于检测某些类型的肿瘤和其他组织畸形。

细胞水平的组织结构也会影响分子扩散。利用扩散磁共振成像技术，可以详细地研究大脑白质（主要由神经胶质细胞和有髓鞘的轴突组成，负责将信号从大脑的一个区域传送到另一个区域）的组织结构，这是因为水分子倾向于朝某些方向扩散，成束的轴突使水分子优先朝着与纤维方向一致的方向扩散。

扩散张量成像利用水在神经组织中的扩散方向受限的特点，生成神经束图像。在扩散张量成像中，每个体素对应一个椭球体，其大小反映了体素扩散的强度和方向。椭球体可以用矩阵（张量）来表征，每个体素的方向信息可用于识别神经束。通过一种叫作纤维束成像的过程，扩散张量成像数据可以形象地表示神经束。通过对单个神经束进行颜色编码，我们有可能得到赏心悦目且信息量大的图像，展现出大脑各区域之间的重要连接。

纤维束成像可用于获取人脑连接模式的大量低分辨率信息。图9－4展示的是利用扩散张量成像技术获得的正中矢状面（将大脑分成左右两半的平面）图像。连接大脑不同区域的轴突纤维，包括连接两个半脑并穿过正中矢状面的胼胝体纤维，在图中清晰可见。

脑电描记术和脑磁描记术在获取脑行为信息时，依据的物理学原理与上述方法不同。脑电描记术和脑磁描记术都是通过观察神经元活动对电磁场的影响，直接测量神经元的活动水平。

当大量神经元群协同放电时，就会产生可被探测到的电场和磁场。最早在动物身上发现并报告这个现象的人是利物浦医生理查德·卡顿。1924年，脑电描记术的发明者汉斯·贝格尔记录下第一张人类脑电图。



图9－4 通过扩散张量成像技术获得的穿过正中矢状面的纤维束图像图片来源：托马斯·舒尔茨。

脑电描记术通过多个电极——通常放置在头皮上——记录大脑在一段时间内的电活动。单个神经元活动产生的电位很低，无法被电极捕捉到。因此，脑电描记术反映的是有相关行为和相似空间定向的许多神经元的关联活动。在头皮上记录的脑电描记术信号，绝大多数是皮质中的锥体神经元产生的，因为这些神经元排列整齐，并且以一种高度相关的方式放电。由于电场与距离的平方成反比，所以探测大脑深层结构的活动难度更大。

脑电描记术信号分析常用于大脑的功能失调性行为的临床检测，其脑电图会呈现异常模式。脑电描记术也被用于大脑研究，但由于信号来源的定位只能精确到几厘米以内，所以它的空间分辨率很低。然而，它的时间分辨率很高（可达到毫秒量级）。脑电描记术有时会与另一种成像技术结合使用，比如磁共振成像或正电子发射断层成像。

皮层电描记术或颅内脑电描记术，是一种与脑电描记术类似的技术，它通过直接放置在大脑裸露表面的电极记录大脑皮质的电活动。皮层电描记术广泛应用于动物研究，但由于这项技术需要去除一块头

骨让大脑表面裸露在外，所以在以健康人为实验对象的基础脑研究中使用得不多。

最早由戴维·科恩提出的脑磁描记术，也是通过探测放电神经元的电磁场效应来发挥作用的，但它使用的探测器是非常敏感的磁强计。为了降低磁背景噪声，科恩使用了磁屏蔽室（图9－5）。

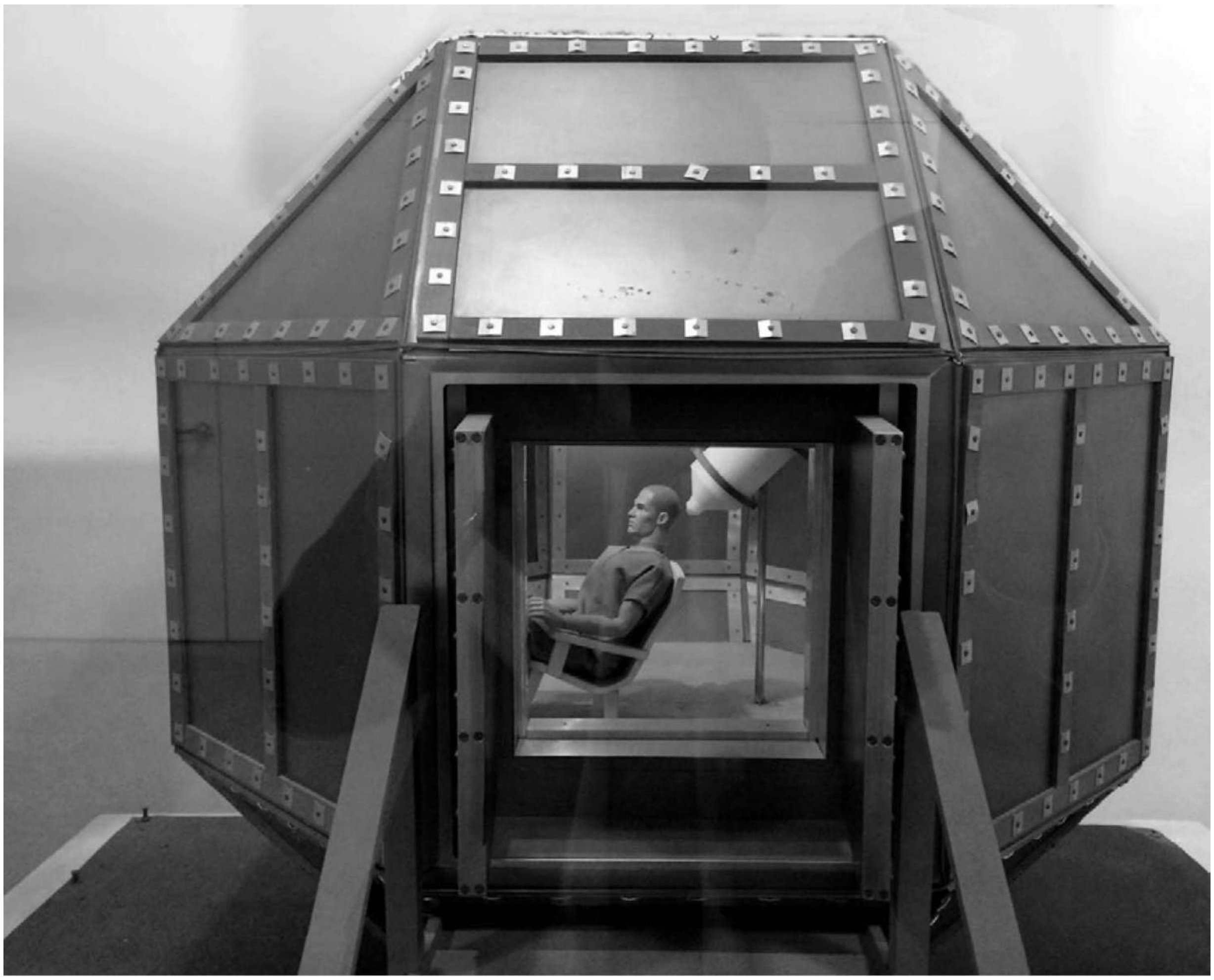


图9－5 1969年，戴维·科恩在麻省理工学院为第一个脑磁描记术实验建造的磁屏蔽室模型
资料来源：摄于艾瑟尼拉－马蒂诺生物医学成像中心。

脑电描记术检测离子流引起的电场变化，而脑磁描记术检测当大量神经元关联放电时神经元内部和外部电流引起的微小磁场变化。由于电流必须有相似的定向才能产生相互增强的磁场，所以（就像脑电描记术一样）垂直于皮质表面的锥体细胞层产生的信号最容易测量。定向与头皮表面相切的神经元束，将其磁场中的很大一部分投射在头

的外部。由于这些神经束通常位于脑沟中，所以脑磁描记术测量这些区域的神经活动的效果比脑电描记术好，而脑电描记术对颅骨附近皮质沟顶部产生的神经活动更敏感。

脑磁描记术和脑电描记术都可以测量由神经元内部电流和穿过神经元细胞壁的电流引起的电磁场扰动，但它们使用的技术大不相同。脑电描记术使用的电极比较便宜，这些电极被放置在头骨上，每一个都连接一个放大器。脑磁描记术使用的是价格昂贵、灵敏度高的磁探测器阵列，即超导量子干涉器件。一台脑磁描记术探测器的安装成本为几百万美元，操作成本也很高。传感器不仅价格不菲，而且必须严格屏蔽外部磁场，还要冷却至超低温。科学家正在努力研发一种无须冷却到超低温的灵敏磁探测器。未来某一天，新技术可能会使磁探测器更接近头骨，从而提升脑磁描记术的空间分辨率。

事件相关光信号是最近研发的一种方法，它依据的事实是，脑组织活动的变化会产生不同的光散射特性。导致脑组织活动发生变化的原因可能是，神经元内部的离子和水运动引起了体积变化，或者是离子浓度的变化改变了神经元内部水的衍射指数。与血氧水平依赖效应不同，当神经元处于活跃状态且空间定位明确时，光的散射就会发生这些变化。事件相关光学信号的空间分辨率仅略低于磁共振成像，但它的时间分辨率要高得多（可达到100毫秒的量级）。不过，事件相关光学信号目前只适用于距离大脑表面不超过几厘米的皮质区域成像。

上述各项技术已广泛应用于研究工作大脑的行为，它们生成的脑活动图谱为大脑过程的宏观研究提供了重要信息。全脑磁共振成像分析帮助研究人员根据神经元在各个体素中起到的作用，将工作大脑中的体素分成了几个类别。大多数研究使用的体素都是立方毫米量级的，接近现有磁共振成像设备的分辨率水平。很多研究旨在整合不同对象的成像数据，并已开发出若干人脑图谱，比如“人脑概率图谱及参考系统”和“大脑图谱”（在人类脑计划的背景下完成的）。

磁共振成像、脑电描记术、正电子发射断层成像、事件相关光学信号等不同技术的集成，可以提升信息收集的质量。目前，这类集成正在如火如荼地进行着。

最近发展起来的光控遗传修饰技术显示出巨大的潜力，可以提升我们对大脑功能的理解。1979年，弗朗西斯·克里克提出，脑科学要取得重大进展，就必须找到一种办法去控制某些脑细胞的行为，同时不能改变其他脑细胞的行为。虽然我们可以用非常细的探针对单个神经元实施电刺激，但这种方法不适用于大量神经元的研究。无论是药物还是在大脑外部产生的电磁信号，都做不到有选择性地瞄准单个神经元，但光控遗传修饰技术有望做到这一点。光控遗传修饰技术利用光控膜通道蛋白质来控制神经元的活动。我们已知这类蛋白质存在于单细胞藻类中，但在21世纪初研究人员报告称，它们可以在培养物和活体动物身上发挥神经元膜通道的作用，并且可用于刺激或抑制单个神经元和小神经元群的活动。利用光刺激特定的神经元群，有助于我们非常详细地研究这些神经元活动对生物行为产生的影响。

光控遗传修饰技术运用基因工程技术将与光敏蛋白质对应的基因插入脑细胞。一旦这些蛋白质出现在经过修饰的模式生物的细胞中，就可以利用适当频率的光来刺激特定区域的神经元。此外，通过开关控制光，对这些神经元活动的控制精度可以达到几毫秒以内。研究活体动物（比如小鼠或苍蝇）时，就可以采用这种方法，通过控制它们大脑的行为，来加深我们对脑回路的理解。

严格地说，光控遗传修饰技术并不是一种成像技术，但它可用于获取活体生物大脑行为的详细信息，它的分辨率是现有的其他成像技术都无法媲美的。

脑网络图谱

前文中介绍的成像技术辅以其他信息，可以帮助研究人员理解大脑的不同区域是如何在结构和功能上相互连接的。它们为我们提供了连接大脑各个区域的轴突束的大量信息，让我们对大脑执行特定任务时不同区域的活动水平有了广泛的了解。

在研究脑网络时，大脑的每个小区域（这是在综合考虑几何结构、生理和功能等因素后划分的）都对应着图像中的一个节点。在进行脑成像并记录各区域的活动水平时，研究对象需根据要求要么执行一项特定任务，要么保持安静状态。随后，研究者利用数据统计分析和图表算法，研究大脑行为和结构的不同方面。如果研究涉及多个研究对象，不同大脑的坐标就会被映射到同一个坐标中。这样做是有必要的，因为大脑的几何结构各不相同，简单、固定的空间坐标无法表示不同大脑的同一个点。因此，研究人员需要利用线性和非线性转换，将各个大脑的坐标映射到某个参考大脑的坐标中，使不同大脑的相同区域有可能建立非常相近的对应关系。

举个例子，我们来看一个依据高分辨率的功能磁共振成像数据建立的大脑网络。艾瑟尼拉－马蒂诺中心和葡萄牙里斯本高等理工学院的研究人员合作进行了一项实验，他们的目标之一是确定大脑区域之间最强的功能性连接。研究人员对9名处于安静状态的健康受试者的大脑进行了高场（7特斯拉）磁共振成像扫描，该扫描仪的各向同性空间分辨率为1.1毫米，时间分辨率为2.5秒。研究人员针对不同的研究对象，选择了若干感兴趣区，然后利用某种方法去识别稳定和显著的感兴趣区。接着，他们依据这些区域活动水平之间的相关性构建出一幅图，并根据相关性的值给出各边的权重。处理图像时，先要确定最大

生成树，然后将感兴趣区之间的其他连接按强度依次递减的顺序添加到生成树中，直到平均节点度达到4。利用第4章介绍找到最小生成树的算法，也可以找到最大生成树。这个过程会生成一个包含124个感兴趣区的网络（图9－6）。感兴趣区的数量越多，图像的信息量就越大，但进行直观解释的难度也越大。

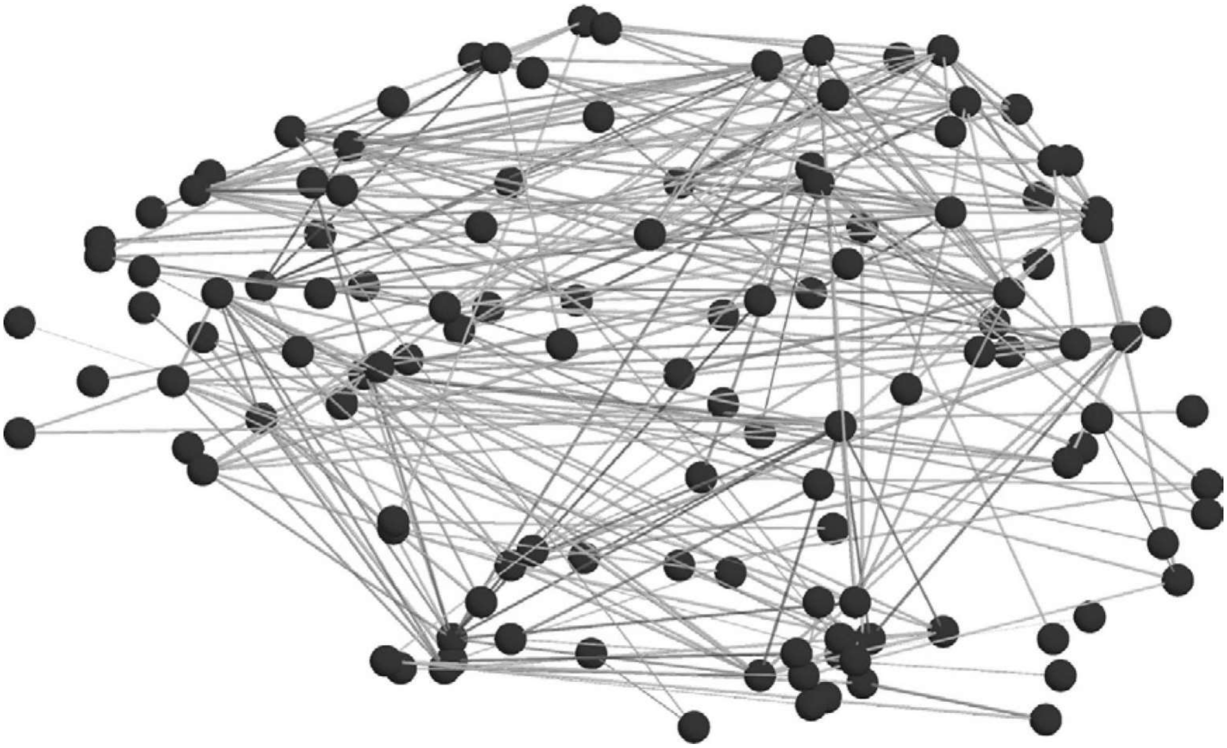


图9－6 依据9名人类受试者的124个感兴趣区之间的最大相关性建立的脑网络

此外，研究人员针对每个受试者的每个成像体素选取一个感兴趣区，由此形成了一些非常大的图像，再利用功能磁共振成像数据对这些图的结构和特征进行研究。这些图像有超过100万个节点和多达10亿条边，具体取决于定义感兴趣区之间的连接时使用的最小相关性水平。

分散在心理学、神经学、遗传学和药理学等不同领域的成千上万项研究，运用了几百种方法研究脑网络。脑网络可用于以下研究：神经回路和神经系统，表征特定活动或特定感觉的放电模式，脑部疾病及大脑行为的许多其他特征，等等。奥拉夫·斯波恩斯写作的《脑网

络》（*Networks of the Brain*）一书涵盖了脑网络研究的多项技术和目标，是相关研究的优秀参考资料。

显微术和脑组织切片

前文中介绍的成像方法的空间分辨率有限，相较之下，我们可以通过在显微镜下细致地观察死亡大脑的切片，轻易地获取关于神经元结构和连接的精细信息。一旦获得了大脑切片，显微镜就可以在不破坏样本的前提下，以足够高的空间分辨率辨识出单个神经元、轴突、树突、突触和其他结构。

传统显微术沿袭了16世纪的显微镜技术：整个样品都笼罩在光下，光程内的所有样品部位被同时激发。反射显微术和透射显微术直接测量反射光或透射光，其分辨率受到若干因素（主要是用于样品成像的光的波长）的限制。可见光的波长为400~700纳米，这意味着人眼无法分辨远小于一微米的特征。神经元细胞主体的尺寸为4~50微米，但突触要小得多，只能填充神经元之间的20~40纳米宽的间隙。因此，只有电子显微术才能分辨出突触等神经元组分。

电子显微术在对样本进行成像时，使用的是电子而不是光子。由于电子的波长只有可见光光子波长的千分之一，所以电子显微镜的分辨率比光学显微镜高得多。电子显微术可以分辨小于0.1纳米的特征，分辨率比光学显微镜高几个数量级。因此，电子显微术已广泛应用于确定单个神经元和小神经元群的具体结构。但是，电子显微术在生物组织中的应用受到些许限制（当应用于活体组织时，难度尤其大）。因此，我们在观察活体组织或者利用与特定生物过程相关的生物标志物染色的组织时，可以采用其他方法。

荧光显微术成像时使用的光不是反射光，而是样本在吸收了光或其他形式的电磁辐射后发出的光。样品必须包含荧光团，即受可见辐射激发后再次发光的荧光化合物。荧光的波长与荧光团吸收辐射的波

长不同，具体取决于荧光团的性质。荧光团发出的荧光可被显微镜的光探测器探测到。用来标记细胞的荧光分子种类不同，并且与不同的生物过程有关，所以荧光显微术可用于跟踪细胞内的特定蛋白质。

共焦显微术使用点照明和靠近样本的针孔，消除来自样本其他部分的离焦光。由于只有非常接近焦平面的荧光发出的光才能被探测到，所以图像的光学分辨率，特别是在深度方向上，要比大视场显微镜好得多。

多光子荧光显微术是光学显微术的另一项延伸技术。就像荧光显微术一样，多光子荧光显微术利用长波长光脉冲激发标本中的荧光团。应用多光子荧光显微术时，荧光团必须吸收几个长波长光子的能量，而且这些光子必须差不多时间到达，这样才能激发电子进入更高的能态。当电子回到基态时，它会发射出光子。通过控制激光源，可以为活体细胞创造出优于其他技术的成像条件。双光子显微术是多光子显微术的一个特例，即荧光团吸收的红外光光子恰好是两个。

无论使用哪种显微术，只有非常薄的脑组织切片才能成像，因为切片深处的部位是无法成像的。用于成像的切片可以是从组织表面切下来的，也可以是组织块的顶层。

为方便切片，人们会对脑组织进行化学硬化处理，之后使用切片机或超薄切片机切片。这些设备简直就是切肉机的升级版，可以切出非常薄的样本切片。切片机可以使用钢、玻璃或金刚石刀片，主要取决于待切片的材料和所需的切片厚度。钢和玻璃刀片能切出适合光学显微术使用的动植物组织切片，金刚石刀片能切出适合高分辨率显微术使用的极薄脑切片。

几十年来，人们一直利用切片机获取大脑切片的二维图像。为了获得关于神经元排布方式的更详细三维信息，研究人员将切片堆叠到一起，然后把成像信息组合起来，以获得完整的三维信息。这种技术被称为序列电子显微术，经过多年的艰苦研究，人们利用它掌握了秀丽隐杆线虫的完整结构。使用序列电子显微术的困难之处在于，当切

片被取出后放入玻璃刀片时，它们会变形，难以对齐。尽管存在这些困难，序列电子显微术仍广泛应用于获取大脑结构的三维信息。

与序列电子显微术相比，温弗里德·登克、海因茨·霍斯特曼提出的序列块表面扫描电子显微术有了显著的改进。这项技术可以对放入电子显微镜的整个组织块进行表面成像。接着，用装有金刚石刀片的切片机从组织块上取下一个薄切片，然后进行切片成像。与序列电子显微术相比，序列块表面扫描电子显微术具有自动化程度高和图像失真小等优点。

序列电子显微术和序列块表面扫描电子显微术得到的图像，可用于显示单个神经元的三维几何结构，甚至是单个突触的位置。从本质上看，这个过程与计算机断层成像相似。二维成像技术和计算机图像合成技术相结合，形成了各种应用于不同环境的技术。通过显微术、组织准备、序列切片分析和计算机图像处理等技术的组合应用，研究人员获得了越来越大的大脑结构的翔实信息。

从概念上讲，完整大脑的整体连接结构图谱是可以绘制出来的，尽管目前的技术水平仍然有限。研究人员在绘制哺乳动物大脑的大片区域的图谱方面付出了大量的努力，他们获得的数据已被纳入开放连接组项目。

2011年，哈佛大学和卡内基－梅隆大学的研究人员利用双光子钙成像（一种用于监测活体脑组织中神经元活动的技术）和序列电子显微术，获得了小鼠视皮质的一定容积（450微米×350微米×52微米，包含大约1 500个神经元）的功能和结构信息。研究人员利用双光子钙成像定位出对特定刺激产生反应的13个神经元，并重建了这13个神经元的连接图像。

2013年，一个研究小组报告称，他们重建了小鼠的一小片视网膜中包含的950个神经元及其相互连接。他们利用序列块表面扫描电子显微术，从一小片面积为114微米×80微米的小鼠视网膜中获取了若干切片。通过人类管理员和机器学习算法对数据进行标注，得出了小鼠视

网膜这一区域的完整结构。据莫里茨·赫尔姆施塔特等人估计，整个标注过程花费了两万多个小时。

2015年的研究报告称，哈佛大学、麻省理工学院、杜克大学和约翰·霍普金斯大学的研究人员利用3纳米×3纳米×30纳米的体素，完整地重建了一个容积为1 500立方微米（比1立方毫米的百万分之一大一点儿）的小鼠新皮质柱中的所有神经元小段和许多亚细胞对象（包括突触和突触小泡）。他们使用装有金刚石刀片的带式收集切片的超薄切片机，获得了2 250个脑切片，每片的厚度约为30纳米。利用序列电子显微术取得切片图像后，通过图像处理完成了若干容积的重建。在这些容积中，研究者重建了大约1 500立方微米的神经组织的三维结构，其中包含几百个树突、超过1 400个神经元轴突和1 700个突触（平均每立方微米约有一个突触）。图9－7转载自描述这项研究成果的一篇论文，显示的是一个新皮质柱的重建效果图。



图9－7 根据纳拉亚南·卡斯瑟里等人的重建结果，绘制的一个直径约为8微米、长度约为20微米的新皮质柱效果图

上述研究和许多其他研究均表明，我们可以获取关于大脑精细结构的详尽信息。终有一天，我们可以获得足够详细可靠的信息，从而确定对脑切片进行神经元水平的精确模拟所需的参数。

脑模拟和超级计算机

随着我们对活体大脑机制和结构理解的不断演化，我们或许可以利用工程学方法复制和详细地研究大脑中的几种不同现象。当我们对大脑组织的精细结构和控制大脑发育的机制有了详细了解之后，就有可能模拟大脑的发育和行为。这些模拟必须复现指导神经元生长的复杂生化机制，突触可塑性（和其他形式的大脑可塑性）的化学和电基础，以及工作大脑包含的几十亿个神经元和几万亿个突触的具体运行情况。不过，如果有了关于这些组分的良好模型，我们在为工作大脑的绝大部分区域建模时，就不会遇到任何根本性障碍。

从计算的角度评估大部分脑部位的模拟乃至全脑仿真，是一件有趣的事。在目前已完成的项目中，蓝色大脑计划可能是最雄心勃勃的项目之一。它是洛桑联邦理工学院与IBM合作开展的一个项目，启动于2006年，最初的目标是在超级计算机上模拟大鼠的一个新皮质柱。新皮质柱可被视为哺乳动物新皮质的最小功能单元，高约为2毫米，直径为0.5毫米；人类的新皮质柱包含大约6万个神经元，大鼠的新皮质柱包含大约1万个神经元。人类皮质可能有多达200万个这样的柱状结构。执行这项任务对计算能力的要求相当高，因为模拟一个神经元所需的计算能力大致相当于一台笔记本电脑的计算能力。不过，超级计算技术正在迅速发展，模拟大多数脑部位将很快成为可能。

在最近进行的研究（其成果于2013年发表）中，为了给大鼠的皮质柱（包含12 000个神经元、500万个树突房室和体细胞房室）建立模型，研究人员在建模和模拟过程中使用了蓝色大脑研究机构的软硬件基础设施。模拟使用的是公开软件，即神经元软件包，运行软件的是一台拥有1 024个节点和4 096个中央处理器的蓝色基因－P超级计算

机。由于目前还没有大鼠皮质柱的详细结构数据，所以研究人员利用神经元连接和突触分布的相关统计信息确定了模型的参数。我们可以推断出模拟结果，进而计算出模拟更复杂的系统乃至完整大脑所需的计算机能力。虽然迈克尔·莱曼等人声称在该模拟中观察到了与现实相符的行为，但并没有确凿的证据表明，模拟大鼠大脑中这个“小”的子系统达到的精细水平适用于更复杂、更强大的模拟。不过，我们暂且认为它确实适用于精确模拟更复杂的系统。

经评估，为完成上述研究而配置的蓝色基因－P超级计算机，它的速度约为每秒14万亿次浮点运算。这使这台计算机能够在3小时的计算机时间内模拟大约4秒的大脑皮质活动，也就是说，模拟系统的用时是实际时间的2 700倍左右。在我撰写本书期间，最快的计算机——拥有1 000万个处理器内核的神威·太湖之光超级计算机——据估计每秒可完成9.3万亿次浮点运算，它的速度大约是上述模拟中使用的蓝色基因－P超级计算机的6 600倍。

由上述模拟推及拥有更多内核的计算机并不是一件易事，但如果我们忽略这方面的问题（这些问题确实有难度，但可以解决），那么从理论上讲，神威·太湖之光超级计算机可以模拟包含12 000个神经元的大鼠皮质柱，而且用时不到实际时间的一半。在假设所有其他条件都相同（尽管我们知道它们可能并不相同）的情况下，如果我们可以建立人脑的完整模型，神威·太湖之光模拟人脑中860亿个细胞的速度可能是实际速度的大约300万分之一。要模拟10秒的真实大脑时间（足以模拟一句较短言语的时间），需要花费大约一年的计算机时间。更令人吃惊的是，超级计算机和真实大脑需要的能量差别巨大。功率约为20瓦的人脑工作10秒钟需要花费大约200焦耳的能量，而15兆瓦的神威·太湖之光在模拟相同事件的一年时间里，需要花费大约5×10^{14}焦耳的能量，是人脑的2万多亿倍。

不过，由于相关技术正在迅速发展，所以模拟变得越来越高效。迄今为止，为测试该技术的极限而开发的最雄心勃勃的脉冲神经元网

络模拟，使用一台超级计算机仿真一个由18亿个神经元组成的网络，每个神经元平均有6 000个突触。这个模型不像蓝色大脑计划使用的模型那么详细，而是以牺牲一些精度为代价来提升模拟速度。

这个模拟由每秒可进行8 000万亿次浮点运算的超级计算机“京”（当时世界上速度第四快的计算机）耗时40分钟完成，目的是模拟脉冲神经元网络1秒钟或实时的行为。若替换成完整大脑的大小，并假设尺度上推不会导致任何负面影响，对人脑中所有神经元的模拟速度仍然只是实际速度的10万分之一。如果所有其他参数保持可比性，那么神威·太湖之光的模拟速度“仅”为实际速度的万分之一。

看到上面这些数字，你的第一反应可能是惊讶、绝望，但也会产生一种如释重负的感觉，因为基于这些可靠的数据，你进一步了解到人脑实在太复杂、太强大、太高效了，计算机根本无法完整地模拟人脑。事实上，即使我们能够建立一个有效的人脑模型（在遥远的将来也许可以），即使专门负责这项任务的地球上速度最快的计算机能够运行这个模型，模拟速度也会比实际速度慢很多。这个简单的推理似乎表明，这样的尝试永远不可能成功。不过，这个看起来显而易见的结论其实过于简单化了，因为它没有考虑到两个不可忽视的因素。

第一个因素是计算机能力的指数增长，这种势头不但没有停止，而且可能会持续相当长时间。如果计算机技术的发展仍然遵循摩尔定律，性能每两年增长一倍，并且这个发展速度也适用于最大的超级计算机，那么到2055年左右，计算机将能实时模拟人脑。这些计算中有若干假设很可能被证明是错误的，包括摩尔定律预测的计算机的持续发展。尽管如此，四五十年后仍然有可能出现这样的计算机。

在我看来，第二个因素甚至比第一个因素更重要。第7章说过，数字芯片的模拟是多级模拟，设备行为的细节在更高层级的模拟中会被抽象化。这意味着，模拟单个晶体管的具体运行情况的多级模型，可以有效模拟包含几十亿个晶体管的系统。多级模拟技术的应用使模拟速度提高了多个数量级，我们没有实质性的理由认为这种方法不能有

效应用于完整大脑的模拟。事实上，多位作者（包括阿多洛拉塔·马拉斯科、亚历山德罗·里蒙吉洛、米歇尔·米格里奥）都提出了加速模拟神经元的方法，它们将使模拟速度得到显著提升。的确，创建一个集成电路的多级模型（明确清晰的模块和易于理解的结构）要比创建一个人脑的多级模型（连接网络复杂，模块不明确）容易得多。但我认为，这一领域未来仍会取得许多进展。随着人们在人脑各部位行为的模拟上投入的努力越来越多，运行这些模型也需要越来越多的资源。我们将创建更复杂的人脑块状结构的多级抽象化模型，并用它们来创建更高级的脑区模型，计算机可以更快地模拟这些模型，而且无须以牺牲模拟工作大脑所需的精度为代价。

在对计算机、生物系统和大脑有了更好的理解之后，现在让我们走进本书的核心问题：我们能够创造出数字思维吗？

第10章

智能、思维和机器

前几章集中讨论了人类大脑的行为，毫无疑问，思维就是这个系统的产物。人类大脑是不是能够产生思维的唯一系统，这个问题值得我们具体讨论。

拥有思维与拥有智能具有一定的相关性，尽管我们有可能想象出没有智能支撑的意识行为和没有思维支撑的智能行为。为便于讨论，假定我们接受图灵测试的前提条件：如果某个系统通过了某个复杂巧妙的图灵测试，我们就认为它是智能的。与图灵测试的复杂程度相关的重要问题，包括相互作用持续多长时间，提问哪些问题，测试人员

必须掌握哪些知识，以及虽然重要但与本次讨论关系不大的诸多细节等，所有这些问题我们稍后会加以讨论。

我们有理由认为，通过图灵测试并被认定为智能的系统，必然有它自己的思维。这个论述会引发一些问题，比如，“拥有思维的意思是什么？”。虽然某个系统从其表现来看像智能系统，并且通过了图灵测试，但该系统可能根本没有它自己的思维，而只是成功地给人一种它有思维的错觉。对于这个问题我们不能掉以轻心。

在本书中，“思维”（mind）一词被使用了很多次，但我们很快就会意识到它的概念是非常模糊的。对于“思维”一词，公认的定义通常是“大脑行为的一种涌现性质，为人类提供了一系列认知能力，包括智能、意识、自由意志、推理、记忆和情感等”。我们可以讨论其他高等动物（比如狗和猴子）是否有它们自己的思维；我们可能愿意承认它们有自己的思维，但它们的思维与我们的思维不同，从某种意义上说，比不上我们的思维。

思维最显著的特征也许是，它为它的主人提供了意识，以及意识能做到的一切事情。毕竟，我们可能会承认在计算机中运行的程序可以推理，可以执行复杂的数学计算，可以战胜人类象棋冠军，可以在游戏类电视节目中击败人类，可以翻译语音，甚至可以驾驶车辆，但我们仍然很难相信具备这些能力的程序拥有自由意志、自我意识和自我的概念，或者担心它们会死去。在这个意义上，我们往往把意识看作人类特有的东西，尽管我们不能准确地解释意识到底是什么，它来自哪里。数以千计的杰出思想家都试图解答这个问题，但答案仍然悬而未决。

经验使我们只熟悉一种特殊的思维，即人类思维，这使得我们很难想象或思考其他类型的思维。当然，我们可以想象有一个计算机程序的表现及其与外部世界交互的方式同人类无异，我们也愿意承认这样一个程序可能拥有思维，但这样的程序目前还没有出现，所以人类和机器之间的差别仍然非常大。

使用图灵测试来检测计算机是否拥有智能，这反映了一种以人类为中心的局限思维观。图灵测试的设计初衷是消除人类中心主义倾向，但它仍然是基于机器对人类行为的模仿。它无法测试一个程序是否真正拥有思维，而只能测试一个程序是否拥有与人类非常相似的思维。

许多思想家、哲学家和作家都讨论过一个问题：非人类思维是否可以从在计算机、计算机网络或其他计算支持体系中运行的程序中涌现出来？但是，如果不能给思维下一个更可靠的定义，我们就无法开发出可以识别一个特定系统是否拥有思维的测试；如果该系统拥有的思维与我们熟悉的人类思维截然不同，这个目标就更不可能实现了。

为了规避上述难题，我们来考虑思维的一个更加严格的定义（我称之为“特定领域思维”）：系统的一种属性，使系统能够在某个特定领域以类似人类的方式智能地运行。例如，一个下棋下得非常好的程序必须拥有下棋的特定领域思维，一个可以在大街上娴熟地驾驶汽车的程序必须拥有驾驶的特定领域思维。当然，特定领域思维受到的限制超过人类思维。不过，考虑到可以在任何复杂领域以类似人类的方式智能地运行的系统的内在复杂性，我们似乎有理由要求读者接受一个观点，即这样的系统必然拥有某种特定领域思维。这种特定领域思维可能比人类思维简单，而且可能无法赋予它的主人意识、自由意志、自我意识，以及任何可能源于自我意识的特征，比如对死亡的恐惧。

特定领域思维是系统的一种涌现性质，它使系统能够充分利用世界的某个内部模型在某个特定领域智能地运行，并且表现得像人类一样聪明和得心应手。不难想象，特定领域智力测试将以人为参照物来测试系统，通过测试的系统会被认定拥有特定领域思维。在这类测试中，图灵测试是最严格的，因为系统不仅要在诸多特定领域中智能地运行，还必须在所有这些领域中表现出与人类十分相近的智能和行为。

显然，有些领域具体而简单，让系统拥有特定领域思维是小事一桩。例如，玩井字游戏从来不会输的系统拥有非常简单的特定领域思维，因为井字游戏很简单，只要将一组简单的规则编写到程序（甚至是一组电线和开关）中，它就可以战无不胜地玩这个游戏了。同样地，如果某个系统可以按照规则下国际象棋，但一局也赢不了，它的象棋思维就非常简单，近似于没有思维。相反，具有国际象棋锦标赛水平并能在大多数情况下打败人类冠军的系统，必定拥有象棋的特定领域思维，仅靠一些简单的规则是做不到这一点的。有了这个附加条件之后，我们可能会承认，即使是像蠕虫这样简单的生物也拥有特定领域思维，使它们能够在特定的环境中生活、竞争，虽然这些思维非常简单，是一个非常简单的大脑运行的产物。

这些定义以某种方式确定了思维的本体和层次性。如果思维A可用于在思维B适用的所有领域甚至其他领域中进行推理活动，我们就可以认为思维A比思维B更具通用性。人类思维可被视为迄今为止已知的最具通用性的思维，它包含多种特定领域思维，并表现出一些产生自我意识和意识的其他特征。这些其他特征是什么，它们来自哪里？后文将讨论这些重要的问题。

现在，我将根据思维的起源、计算支持及其展现出的智能形式，对思维进行分类。如果思维是设计的产物，而不是通过进化“自然”出现的，我将称之为合成思维。如果思维是通过进化产生的，我将称之为自然思维，比如人类思维，其他动物的思维可能也属于这个类别。如果思维是由数字计算机程序运行产生的，我将称之为数字思维。如果思维是由生物大脑运转产生的，我将称之为生物思维。图10－1展示的就是这几种类型的思维。

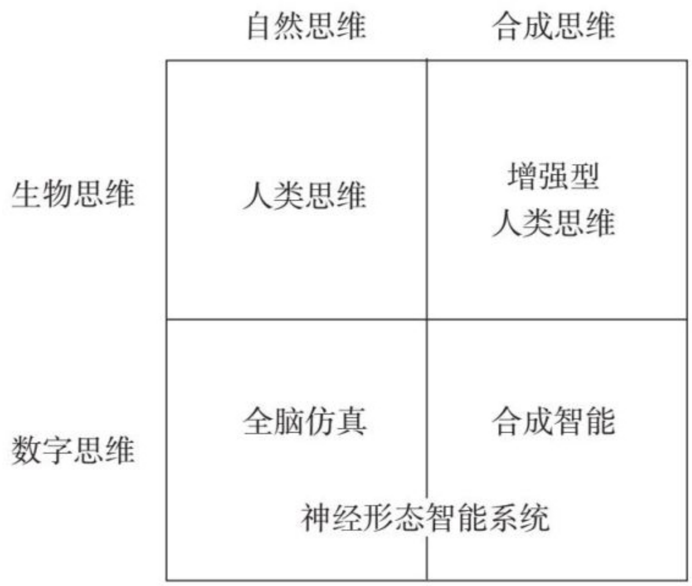


图10－1 数字思维、生物思维、自然思维和合成思维

我们都很熟悉进化设计的自然思维。合成思维是由进化以外的过程设计的，旨在满足特定的需求，在未来几十年内这种思维可能会有所发展。它们也许会使用数字支持系统，但如果在发展过程中使用了合成生物学工具，那么从理论上讲，它们可以靠生物系统来支撑。

图10－1列出的各种思维并非同样直观。几乎每个人都熟悉自然和生物思维，其中最具代表性的是人类思维。许多人也熟悉合成和数字思维的概念，它们的设计原理与大脑运行方式没有任何关系。（“合成智能”属于后一类思维。）我还讨论过未来是否有可能通过在数字计算机中模拟大脑发育过程的原理来获得智能系统，我把这样的系统称为“神经形态智能系统”。此外，将来某一天我们还有可能掌握可以事无巨细地模拟或仿真人类大脑行为的技术，而不是受大脑工作原理启发设计出来的系统。这种方法被称为“全脑仿真”，相当于在数字思维基础上形成的自然思维（因为这种思维的运行方式与进化设计的人类思维完全相同）。

我们还有可能在生物思维的基础上创造出合成思维。终有一天，人类或许可以利用合成生物学技术和先进的遗传工程，设计出具有超人思维能力的生物。

有了思维本体，我们现在就可以尝试回答本书的核心问题：在未来的几十年里，从数字计算机中运行的智能程序中涌现出来的会是哪种类型的数字思维，是特定领域思维还是通用性的自然思维或合成思维呢？

合成智能

现在的系统都是为了展现复杂的特定领域思维而创建的，所以易于辨识。它们通过获取大量信息并使用复杂的算法来执行任务，这显然需要先进的特定领域思维。这种由大型程序员团队开发的系统执行的复杂任务，如果交由人类来执行，则会需要巨大的智能投入。我认为它们是通用合成思维的前身，总有一天会得到人们的认可。随着这些系统的进化，它们可能会变得越来越智能，直到最终成为合成智能。

深蓝是IBM公司开发的一台超级国际象棋计算机，它的前身——深思是由卡内基－梅隆大学开发的。在接受IBM的雇用之后，深思的创造者面临的挑战是，建造出一台能够打败人类世界冠军的国际象棋计算机。1996年，深蓝在第一局比赛中击败了世界冠军加里·卡斯帕罗夫。但到整场比赛结束时，卡斯帕罗夫以4：2的总比分击败了深蓝。之后，IBM对深蓝做了一些改进。1997年，它再次与卡斯帕罗夫对弈，并成为第一个在标准国际象棋比赛中击败卫冕世界冠军的计算机系统。卡斯帕罗夫一直不接受他的这次失败，他说，他有时会从深蓝的一些走法中看到深不可测的智能和创造力，这表明有人类棋手在为这台机器提供帮助。IBM始终否认有任何作弊行为，并指出规则允许深蓝的开发人员在比赛休息期间“调整”机器的程序，而他们只是充分利用了这个机会。卡斯帕罗夫要求重赛，但并未成行——IBM终止了深蓝项目。当然，深蓝是否真比卫冕国际象棋冠军棋艺更高，这一点儿也不重要。在曾被认为必须有强大的人工智能才能获胜的对弈中，深蓝和许多其他现存的国际象棋计算机都有可能击败99.9999%的人类。深

蓝下棋的方式在很多方面都与人类不同，但不可否认的是，深蓝拥有下棋的特定领域思维。

有些游戏比国际象棋更复杂。起源于2 500多年前的中国围棋，它的棋盘上有19×19个位置（图10－2）。玩家轮流在棋盘上放置黑子和白子，以便控制领地。棋子落到棋盘上之后，便不能移动了，这就是所谓的“落子无悔”。每一步都有几百种可能的走法。围棋对计算机来说难度很大，因为它的分支因子很大，搜索树很深。直到2015年还没有任何围棋程序能打败一个优秀的人类棋手，而且人们相信围棋将在很长一段时间内超出计算机的能力范围。但是，2016年，谷歌旗下的深度思维（Deepmind）公司的一个研究团队利用经过深度学习训练的神经网络，创建了一个名叫阿尔法围棋（AlphaGo）的程序，并让它参加了最高水平的围棋比赛。2016年1月，阿尔法围棋击败了欧洲冠军樊麾。两个月后，它击败了世界一流棋手李世石。

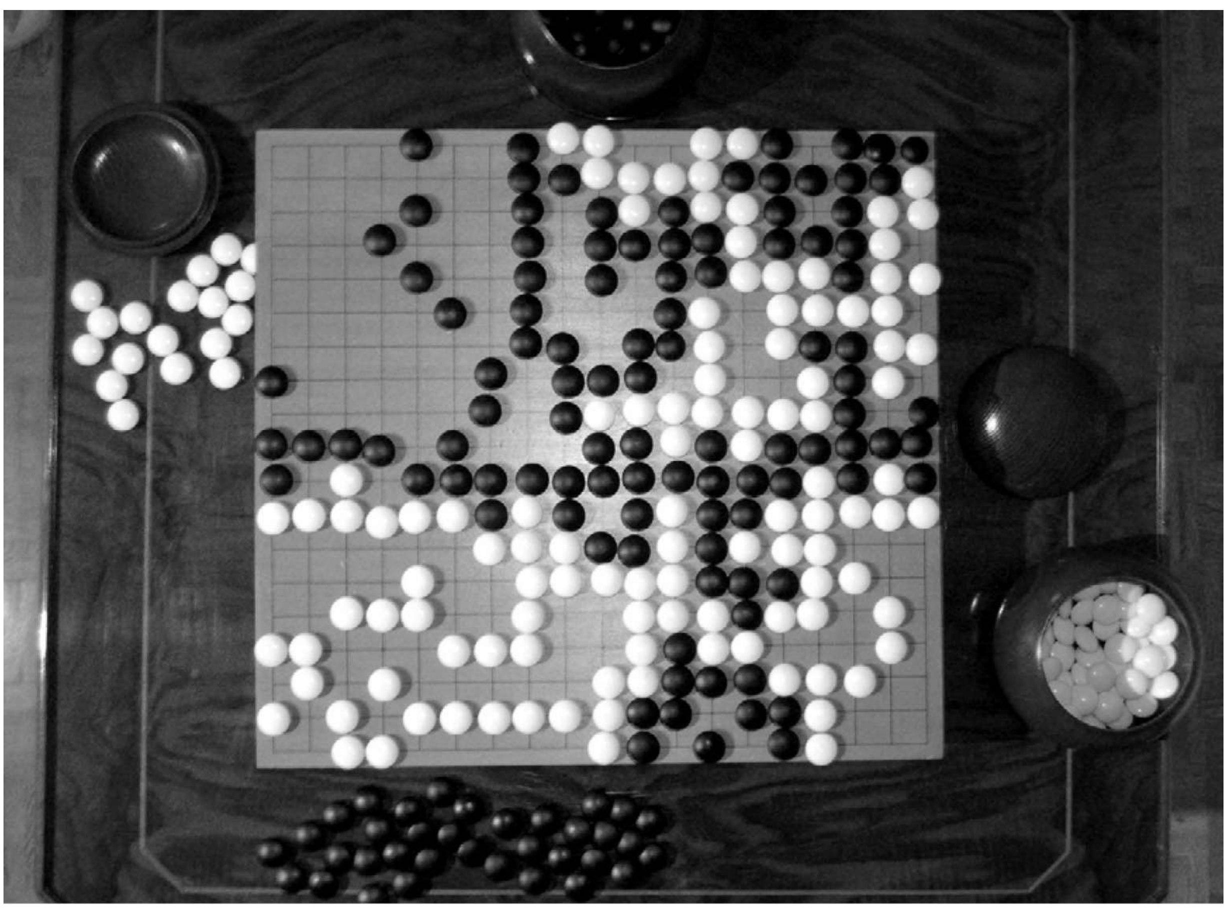


图10－2 一局围棋比赛的最终局面，对弈双方的领地已定

计算机在下国际象棋、下围棋或玩类似游戏时，通常会先搜索并尝试多种走法，然后评估这些走法可能会产生的几百万种局面。在国际象棋或围棋中，局面评估的难度很大，但依据棋盘上棋子的数量和它们的位置，就可以很好地估计局面分值，特别是在有大型棋局数据库可查的情况下。虽然我们无法确切了解人类是如何玩这类游戏的，但我们知道计算机使用的方法主要基于蛮力搜索，在选择一种走法之前会尝试多种可能的走法。计算机已经完全解决了下跳棋的问题，也就是说，所有可能的走法组合都尝试过，并被录入数据库。在任何情况下，只要访问数据库，计算机程序就能找出最好的走法。这意味着，在不访问数据库的情况下，没有人能打败这样一台计算机。（如果可以访问数据库，人类的棋艺也能臻至完美，但任何人的记忆都不可能存储海量的数据。）

但对于比跳棋更复杂的游戏，比如国际象棋和围棋，也许永远无法找到精确的解决办法，因为可能的棋局数量真的太大了。尽管我们还不知道国际象棋的可能棋局的确切数量，但估计超过10^{120}，这个数字远大于宇宙中的原子数量。围棋的可能棋局数则超过10^{800}。因此，我们不能认为深蓝或阿尔法狗下棋靠的是蛮力（也就是说，它们知道所有可能的局势）。深蓝肯定有国际象棋思维，阿尔法围棋肯定有围棋思维，但它们的思维与人类棋手的思维不同。

人工智能的应用之一尽管不太明显，但几乎每天都被很多人使用。在我撰写本书时，万维网上有几百亿个网页，搜索引擎谷歌可以让每个人访问其中大约500亿个网页。通过使用索引技术（像电话簿一样，将包含特定项的网页罗列在一起），谷歌可以在几分之一秒的时间内检索出与用户所选主题相关的网页。此外，谷歌还能将最相关的网页排到前面。为了实现这一点，谷歌在其创始人谢尔盖·布林和拉里·佩吉提出的网页排序算法的基础上做了一些改进，采用了一些聪明的方法。

网页排序算法是谷歌成功的根源，它将网络视为一个数量巨大、相互连接的网页集，并且认为如果用户通过连接页面的超链接随机浏览这些网页，访问频率更高的页面很有可能就是他们感兴趣的网页。因此，如果有更多的链接指向某些网页和域，我们就可以认为它们更吸引人。这也是为什么我们更有可能检索CNN（美国有限电视新闻网）或BBC（英国广播公司）的页面，而不是某个更模糊的新闻来源。相较于不太知名的网页，指向CNN（为了获取新闻或出于其他原因）的网页更多。多年来通过其他多项技术的补充，网页排序算法可以对查询结果网页进行良好的排序。

通过检索这一非常简单的操作，我们可以得到一组与查询项相关的页面，这得益于计算机科学领域的工程师和研究人员多年来开发的十分先进的技术。很多人常用谷歌检索模糊主题的相关信息，但我们可能已经忘记了，仅仅在20年前，这样的操作即便可行，速度也非常慢。

我们通常认为谷歌没有它自己的思维，甚至没有特定领域思维。现在，假设我们回到了1985年。就在几年前，人们还在开发和部署能让计算机之间进行远距离通信的TCP/IP协议。而到了1985年，人们已经可以使用计算机－计算机网络发送和接收消息（主要是电子邮件和文件）了。但只有几千人在使用计算机的这种功能，而且没人能预测到在短短几十年后，将会有几十亿台计算机和移动设备像现在这样相互连接。不过，基本技术在当时已经存在，计算机也可以相互通信了。只要有更先进的算法和程序、更快的计算机、更便捷的互连网络，就可以建成我们今天拥有的互联网。

现在想象你回到了1985年，有人告诉你，30年后计算机系统能够在几分之一秒的时间内，找到世界上与你选定主题相关的所有信息。他还会告诉你，这样一个系统可以检索互联网上的所有信息（全世界存储在所有互连计算机上的信息大约为10^{24}字节），在充分理解这些信息的基础上选择与你的请求相关的信息，并在几分之一秒的时间内

向你提交答案。为了厘清头绪，你可以记住一本书通常包含50万字节，美国国会图书馆藏有大约1.5亿本图书，因此互联网上的可用信息相当于100亿座国会图书馆存储的信息。（公平地说，互联网上存储的数据绝大多数都是视频和照片，而搜索引擎只覆盖了其中一部分，因为它们为视频和照片编制索引的主要依据是关键词和标签，而不是内容。不过，随着计算机处理图像的能力越来越强，这种情况很快就会发生变化，但这与接下来要提出的问题关系不大。）

现在，假设你回到了1985年，有人问你：这样的系统是智能的吗？它拥有自己的思维吗？搜索海量信息，并从中立即检索出相关数据，是否需要智能？系统是否必须充分理解文档中书写的内容，才能检索出相关信息并丢弃无关信息呢？

我的猜测是，如果回到1985年，你的回答可能是：这样一个系统肯定是智能的，而且拥有它自己的思维，至少在一定程度上是这样的。毕竟，想要回答它接受的复杂查询，它就必须拥有某种世界模型，并且至少能做到部分理解查询内容。

如果那时的你不知道查询是如何被提交给像我们现在拥有的这种系统的，你可能会想象它是通过某种书面或口头对话完成的。比如，用户输入某个问题（“美国国会图书馆有多大？”），并得到“美国国会图书馆是世界上最大的图书馆，书架全长约838英里，藏书超过1.58亿册”之类的答案。今天，如果你输入同样的问题，那么你得到的文档中既包括你正在寻找的信息，也有一些额外信息——可能与你有关，也可能与你无关。你或许已经非常习惯于使用搜索引擎，以至于你根本没想过它可能拥有自己的思维，但在1985年，你可能会推测它拥有自己的思维。之所以会出现这种标准二元性的情况，是有其原因的。

原因之一是，搜索引擎的用户界面远未臻至完美，与按照用户需求准确检索内容的目标相去甚远。我们主要通过文本与搜索引擎会话（尽管现在我们也可以使用不太完美的语音界面），搜索引擎也不会

给我们写出一个明确的答案，而是提供一个文档列表，我们可以从中找到查询问题的答案。将这样的答案转换成用简单的话语表述的答案，技术条件可能已经具备。事实上，其他系统（比如，IBM的沃森）已被证明具有这种能力。

沃森利用先进的算法处理自然语言，进行自动推理，并通过机器学习回答特定领域的问题。凭借来自维基百科和其他知识源（包括百科全书、字典、文章和书籍等）的信息，沃森可以在游戏类电视节目《危险边缘》中与人类冠军同台竞技，并多次击败优秀的人类对手。（在这个游戏中，通常意义上的问题作为答案呈现在选手面前，而选手需要回答出相关问题。为便于讨论，在本书中，前者将被称为问题，后者则被称为答案。）

观看《危险边缘》节目的视频（可以在YouTube视频网站上找到），会让人产生一种谦卑感和清醒感。计算机运用复杂但易于理解的技术，解释用自然语言提出的难题，并且能比老练的人类玩家更快、更准确地解答这些问题。沃森的电脑合成语音以惊人的精确率回答难题，展现出世界上少有人拥有的文化和历史知识储备。

不过，持怀疑态度的人很快就指出，沃森并不理解那些问题和答案，当然也不知道它赢得了一场比赛。其中最直言不讳的怀疑论者之一是约翰·塞尔，他的论据就是中文房间实验的基本理念。塞尔在2011年发表的一篇文章中指出，沃森做的不过是按照程序规定的一系列特定规则处理符号。他认为，符号处理与理解根本不是一回事，所以沃森不理解任何问题或答案。对程序来说，问题和答案都是无意义的符号。系统无法理解这些问题，因为它无法从符号中获取意义。你可能会问，“那么人脑是如何做到这一点的呢？”塞尔的回答是，人类大脑通过简单的运行，就会产生意识、理解，以及随意识和理解而来的一切。

塞尔和其他怀疑论者认为沃森不理解它听到的问题，也不理解它给出的答案，但我并不赞同他们的观点。我认为理解不是一个神奇的

过程，就像开灯或关灯那样。理解是一个循序渐进的事物，它源于创建一个与当前问题相关的内部世界模型。一个没有理解能力的系统，是无法像沃森那样准确地回答《危险边缘》游戏中提出的复杂问题的。沃森思维中的世界模型足以让它准确地回答绝大多数问题。有些问题它完全不理解，有些问题它只是部分理解，还有很多问题——可能是大多数问题——它完全理解。

因此，问题不在于沃森是否拥有智能。当然，它拥有足够的智能回答许多复杂的问题，既然它能做到这一点，那么它一定有它自己的思维——专门用于《危险边缘》游戏的特定领域思维。沃森知道它赢得比赛了吗？或许知道，因为它的行动目标就是赢得比赛。

沃森会为自己赢得比赛而高兴吗？它意识到自己赢了吗？可能没有，因为沃森要想意识到自己获胜，就需要对游戏、世界和它在世界中扮演的角色有更深刻的理解。据我所知，沃森并没有这些概念的内部模型，也没有自我意识，因为内部模型和自我意识都没有囊括在它的设计里。

沃森已应用于医疗法诊断和石油勘探领域，在这些地方，沃森的答题能力有了用武之地。它可以回答用自然语言提出的问题，加上它拥有以证据为基础的学习能力，可用作临床决策支持系统。它也可以帮助医生为患者找到最好的治疗方案。一旦医生描述出症状和其他因素并提交系统查询，沃森就会分析这些输入，查找最重要的信息，然后找出与当前问题相关的事实，包括数据来源、患者的既往病史和遗传病史。在石油勘探领域，沃森可用作地质学家和数值算法之间的交互界面，算法则负责执行复杂的信号处理任务，以确定潜在地下油田的结构。

深蓝、阿尔法围棋、谷歌和沃森都是以自上而下的方式设计的系统，拥有特定领域思维。深蓝和阿尔法围棋擅长复杂的棋类游戏，谷歌可以响应查询从网络上检索信息，沃森能用自然语言回答特定领域的问题。未来10年，我们可能还会设计和部署其他系统：仅凭世界模

型而无须人类辅助就可以在大街小巷中穿行的自动驾驶汽车，可以帮助我们管理日常事务的智能个人助理，可以用简单易懂的语言与客户交谈的销售代理，等等。到目前为止，所有这些能力都是人类独有的，但在未来10年内它们肯定会被机器掌握。

为解决特定问题而设计的系统，比如上面提到的那些，可能具有足够先进的特定领域思维，可以在特定领域参与对话，但它们不太可能具备形成通用思维的性质。但是，如果把多种特定领域思维整合起来，就有可能形成通用性更强的系统。未来20年，或许会出现计算机化的个人助理，它们可以熟练地完成多种任务，比如，回答问题，从网上检索信息，管理个人日程，安排会议，采购与准备家庭用品，等等。完成这些任务所需的技术在未来20年内肯定会出现，对这种个人助理的需求也会出现（也许可以通过手机或笔记本电脑的界面操作）。

这些系统将拥有更具一般性的思维，但它们仍将被限制在少数特定领域的环境中。为了有效地执行任务，它们必须熟悉人类生活的许多细节；为了对各种要求做出有效的响应，它们必须记住过去的事件。

我预期我们都会认为这样一个系统拥有它自己的思维，可能是一种受限的和多少有些简化的思维。我们也许仍然认为它没有意识、个性、情感和欲望。不过，这些特征很可能主要取决于旁观者的看法。即使是模仿动物行为和情感的简单系统也能激发出情感和意识知觉，当我们阅读泰雷尔·麦尔丹那于1978年出版的小说《安娜·克莱恩的灵魂》（收录在侯世达和丹尼特的《心我论》中）对马克三世怪兽的描述时就能明显体会到这一点。如果对沃森或谷歌进行修饰，使其与用户的情感互动更加深入，我们对它们的个性、情感和欲望的感知很可能就会发生颠覆性的变化。

凯文·凯利在他的颇具见解的著作《科技想要什么》中，用“科技体”（technium）一词来命名由计算机、程序、信息、设备、工

具、城市、车辆和世界上已经存在的其他技术工具构成的巨大互连网络。凯利认为，科技体已经是一个包含几十亿思维的集群，我们之所以没有意识到这一点，只是因为我们有一种“沙文主义偏见”，拒绝承认除人类以外的其他东西也拥有思维。

对凯利来说，科技体是复杂进化过程的不可避免的下一步，地球上的进化已经进行了40多亿年，宇宙中的进化时间更长，早在地球诞生之前就已经开始了。他认为，科技体也是一个自主的实体，有它自己的日程和未来。这是一种培养越来越强大的思维的新方法，过程中会创造出更大的复杂性，开启更多的可能性。我们没有意识到互联网是一个始终存在的思维实体，拥有几十亿个眼睛、耳朵和大脑，这很可能是因为我们闭目塞听，或者是沙文主义思想作祟，或者两者兼有。对于我们创造和发展的合成思维，我不知道我们的看法是否会改变。

神经形态智能系统

智能系统的设计另有他法。一种方法是从人类大脑中汲取更多灵感，人脑系统支持着我们知道的唯一通用思维。以这种方式设计的智能系统将会引发有关思维是什么，以及合成思维和自然思维之间有何区别等更加复杂的问题。如果我们以人类思维运行的基本原理为基础设计智能系统，那么我们可能更愿意承认智能系统有它们自己的思维和意识。毕竟，如果我们想创造一个有意识的智能系统，最显而易见的方法就是复制我们已知的有意识的智能系统，即人类思维。有两种方法可以完成这种复制。

一种方法是直接复制人类思维，我们可以通过巨细靡遗地复制人脑来实现这个目的。人脑是支持人类思维的物理系统，复制工作大脑的所有细节将是一项非常艰巨的工作，而一旦取得成功，就肯定会带给我们预期的结果。毕竟，如果生成系统完整地复制了原始系统并可以正常工作，它的行为与初始系统也没有任何区别，那么它肯定是智能的。这样的思维是现有生物思维的完美拷贝，所以它的行为与自然思维一模一样，尽管是数字计算机在为它提供支持。这种思维设计过程通常被称为思维上传或全脑仿真。

设计类人思维的另一种方法是复制人脑的创建机制，而不是工作大脑的具体细节。这种方法的难度可能没那么大，而且在不久的将来技术的发展可能会使这个方法更可行。它的基本理念不是在计算机上复制工作大脑的细节，而是复制人脑结构的创建原理和机制，并利用它们来创造合成大脑。

第8章介绍了在大脑的自我组织过程中发挥作用的两个主要因素：基因编码和大脑可塑性。随着我们对这两种机制的理解不断加深，终

有一天我们可以在程序中复制大脑的自我组织方式。

许多研究人员认为，像大脑这样的复杂系统并非源于非常复杂的配比（精确调整的参数和详细的连接指令），而是源于某种通用机制，根据在广泛的参数范围内有效的通用规则，从简单的局部交互创建出复杂系统。从大脑尤其是皮质的放电模式中获取的重要证据证明，大脑通过自我组织使自己处于某种临界状态，以确保放电量正好可以充分传递信息，同时避免活性衰减和网络激发失控。

了解大脑自我组织的通用数学原理，以及控制大脑发育的生物和生化机制，最终可以帮助我们在计算机上模拟大脑的发育过程。模拟结果必须与我们从简单和复杂生物处获得的数据进行广泛的对比。如果成功，这类模拟应该可以生成连接结构和细胞活动的模式，并且在所有可观测的方面都与真实的大脑相同。如果物理学和数学原理是正确的，可用于调整参数的知识也是充足的，我们就有理由预期生成的自组织脑在大多数方面（即使不是所有方面）的行为都与真实的大脑相似。这样的模拟无须在神经元的层级上进行，因为更高层级的结构可能参与了大脑组织的过程。事实上，模拟过程甚至无须考虑神经元，因为神经元可能不是合适的抽象化层级。

当然，如果无法全面接收各种刺激，类人脑就不能正常发育。这些刺激只能通过模拟感官输入系统，在适当的大脑结构的发育过程中起着不可或缺的作用。通过这种路径设计大脑的整个过程耗时较长，因为形成大脑组织的机制需要多年的时间去发展。在这个过程中，与现实世界的交互是一个限制因素。从理论上看，整个过程无疑比在真实大脑中的运行速度快很多倍。事实上，第一个方法有可能利用这种可能性，使我们在短短几天内模拟出几个月或几年的大脑发育过程。不过，随着模拟变得更加精确和复杂，我们有必要为模拟脑提供现实的输入模式，使它可以与物理世界实时交互。如果我们的目标是通过这一路径模拟幼猴的工作大脑，这个模拟系统就必须与其他猴子互动，无论是真实的还是模拟的。如果我们的目标是模拟人脑的完整发

育过程，模拟系统就必须与人类和其他主体互动。这意味着，整个模拟系统在大多数时间里必须实时运行，模拟幼儿的发育过程需要几年时间，模拟年轻人的发育过程则需要20年左右的时间。

人脑可以基于第一原理、实验数据和真实世界的刺激来“培育”或“养成”，这种可能性引发了很多复杂的问题。如果这种方法最终会生成一个能思考、懂语言、可以感知世界并与人类交互的系统，那么它也许可以出色地通过任何一种无限制条件的图灵测试。即使交互总是通过模拟世界（从未真实存在的世界）进行的，这一事实也不会削弱智能和意识系统存在的真实性，以及它的情感的真实性。

由此形成的是一种合成思维，但它的工作方式是受自然思维工作方式启发的产物，所以它在某种程度上是自然思维和合成思维的杂交体（图10－1）。如果这样一个系统被开发出来，我们就必须与第一种在某些方面与我们相似的思维友好相处，尽管它没有身体，也不是由人类繁衍的。它的大脑可以像人类一样做计算，它拥有感觉，也拥有在一个只存在于计算机存储器中的世界里生活的记忆。它的感觉可能和我们一样真实，情感可能和我们一样生动，生存欲望也可能和我们一样强烈。我们将模拟人类的自然思维，第一次创造出合成思维。

全脑仿真

前面说过，还有一种方法可以创造类人思维，那就是直接复制活人的工作大脑。这种方法生成的是数字思维，也是自然思维，因为它的工作方式将与人类的生物思维完全相同。不过，为这种思维提供支持的是在计算机中运行的脑仿真器，而不是生物大脑的活动。到目前为止，这是最雄心勃勃也是技术难度最大的方法，因为它需要的技术还不存在，甚至可能永远不会存在。它还提出了一系列极其复杂的问题，因为它将彻底改变智能是什么及其对人类意味着什么等问题的范式。

要依据现有的人类思维创造出数字思维，就必须对工作大脑实施非常详细的逆向工程。从概念上讲，有很多可能的方法，但这些方法都需要了解关于特定人脑的结构与神经活动模式的翔实信息，即可以在计算机上仿真大脑行为的精确可靠的信息。这种方法被称为全脑仿真或思维上传，其基本理念是，基于详细信息的全脑模拟是可以实现的。包括雷·库兹韦尔和尼克·博斯特罗姆在内的多位作者都曾提出，全脑仿真有朝一日必然实现，但直到最近，我们才开始了解实践这样的壮举需要付出什么代价。

从理论上说，思维上传可以通过多种方法实现，但有两种方法尤其值得关注。一种方法需要大规模地复制神经结构，另一种方法需要逐步置换神经元。第一种方法通过扫描和识别大脑中的相关结构，然后将这些信息复制并存储在计算机系统中，计算机系统再使用这些信息进行大脑仿真，从而实现思维上传。第二种方法更加循序渐进，原始大脑将继续运行，但它的计算支持将逐步被置换。

第一种方法，即大规模复制工作大脑，需要满足的主要条件包括：对现有技术进行大幅改进，同时开发出一些新技术。我在第9章介绍了一些可用来识别人类大脑详细结构的技术，在第8章介绍了一些利用神经元的精细模型、相互连接模式和其他相关生化信息，模拟脑组织切片的技术。原则上，如果已知每个神经元和神经元之间相互作用的详细信息，我们就可以利用现有的脑组织结构知识，非常准确地模拟人脑（尽管速度缓慢）。挑战在于，我们目前还没有足够先进的技术可用于获取脑仿真器所需详细水平的结构信息。

现在，许多先进的成像方法都可用于检索大脑结构和神经元活动的相关信息。在第9章，我们探讨了一些可用于获取脑功能信息（大脑的连接方式，执行具体任务时神经元的活动水平等）的技术，包括正电子发射断层成像和磁共振成像。不过，即使是当今最强大的技术，在不侵入的情况下，也只能获取不小于1立方毫米的结构信息，因为1立方毫米是它们能分辨的最小体素的大小。1立方毫米的皮层中有5万~10万个神经元，每个神经元都有几百个或几千个连接其他神经元的突触，因此每立方毫米皮质中的突触总数可以达到10亿量级，每个突触的大小为20~200纳米。

因此，目前的成像技术无法获得大脑中突触位置和神经元连接的详细信息。磁共振成像的实际分辨率取决于所用的具体技术，但有报告称可实现高达每立方毫米10亿体素的分辨率，即体素的边长约为1微米。不过，要达到如此高的分辨率，还需要克服许多障碍，我们也不清楚这种技术是否可用于为工作大脑中的大片区域成像。

要分辨单个突触，就需要利用显微术，这类技术现在可使脑组织切片成像达到前所未有的详细程度。但是，它们会损坏组织，所以只能用于死亡的大脑。我们在第7章中看到，人们费尽力气才取得秀丽隐杆线虫的大脑切片，并用电子显微术拍摄下来，最终获得了它的302个神经元的完整连接示意图，整个过程耗时12年。如果将这样的做法应

用于人脑，哪怕只考虑神经元和突触数量的尺度上推对应的工作量，也需要耗费与宇宙年龄相当的时间。

不过，技术仍在不断发展，也许有一天我们可以对完整的人脑实施类似的过程。在关于秀丽隐杆线虫的研究中，这个过程主要是手工完成的，而现在几乎完全可以实现自动化，由机器人和计算机负责大脑切片、电子显微术和神经元的三维重建等工作。正如第9章指出的那样，小容积新皮质的重建已经达到了非常精细的水平。有了合适的技术，从概念上讲，这样的过程也可以应用于完整的人脑，从而非常准确地复制原初人脑（遗憾的是，它会在这个过程中遭到破坏）。以这种方式收集的信息，其精细水平可能不足以准确地仿真被切片和成像的大脑，因为即使在这个分辨率水平上，影响神经元行为的许多参数仍然存在显著的不确定性。但是，结合大脑的实时行为数据（当然是在大脑正常运行时获得的），我们或许可以将缺失的信息填补上。我在第9章简单提及的光控遗传修饰技术，可用于提供关于大脑具体行为的非常有用的信息。

利用现有技术显然无法轻松地获得关于工作大脑的足够详细的信息，也就无法满足脑仿真器生成足够准确输出的输入需求。从大脑中获取如此详细的信息，显然是现有技术力不能及的。利用显微术获得的脑结构信息非常详细，至少原则上可以为脑仿真器提供输入，尽管很多信息只能通过间接方式获取。但现有技术的局限性导致它们只能应用于几分之一立方毫米的小容积，不过我们可以预期，借助更先进的算法和更优良的显微术，在未来几十年内我们有可能非常精细地完成小鼠全脑的逆向工程。

然而，我们还不清楚，利用现有的显微术（比如序列块表面扫描电子显微术）获得的结构信息是否足够详细，将其输入脑仿真器后是否可以忠实地复制原初系统。难点在于，系统的行为很可能对若干参数敏感，而这些参数很难根据纯粹的结构信息（比如，细胞膜和突触的性质），以及神经元外部环境中的化学信息准确地估算出来。

理想情况下，详细的结构信息应该与活体的功能信息整合起来使用。有了这些信息的整合，我们就有可能通过调整仿真器参数，在计算机上复现原初系统的行为。虽然人们正在开发将这两类信息结合起来的多种方法，但如何将活体大脑的功能信息与显微术扫描脑切片获得的详细结构信息整合起来，仍然是一个挑战。所以，目前尚不可能利用关于单个神经元和神经元群行为的大量知识去构建一个完整的大脑模型，因为我们还缺少关于神经元结构、性质和连接的必要信息。

事实上，我们对不同大脑过程的理解水平参差不齐。一方面，许多研究项目正在收集和组织大量宏观层次的资料。这些信息可用于确定大脑的哪些区域参与了特定的过程，但不能用于详细了解单个神经元或小神经元群的行为。不过，它们可以让我们对大脑在包含几百万神经元的宏观结构层次上的自我组织方式有一定的了解。另一方面，研究人员已经非常详细地了解了在单个神经元中发生的诸多小规模的过程和结构，包括突触可塑性和膜放电行为。

没有显见的方法可以填补两者之间的巨大沟壑。由于物理限制，现有的技术本身不太可能确定所有的必需信息，如果没有这些信息，就无法构建出精确的脑仿真器。纳米技术和其他技术或许可以填补这个鸿沟，但在技术成熟到足以保证结果之前，它们还不能应用于健康的人脑。即使这些技术发展成熟，它们的应用也会引发重大的道德问题。利用动物模型的那些正在进行的项目，包括开放蠕虫项目、开放连接组计划和人类脑计划，有助于我们了解其他方法的可行性，但寻找更精确的方法仍然是未来几十年众多项目关注的焦点。

我无法告诉你如何应对这个挑战，因为我也无能为力。世界各地的多个研究团队正在想方设法寻找可以帮助我们更好地理解大脑的技术。但到目前为止，没有哪项技术或哪个技术组合有望为全脑仿真提供可用的详细信息。有人甚至倾向于认为，考虑到物理限制和这项挑战的复杂性，我们可能永远无法获得构建脑仿真器所需的足够详细的信息。

要使思维上传成为现实，许多技术必须取得巨大的发展，包括先进的显微术、脑组织三维重建所需的计算技术，以及多级神经模拟和仿真技术。处理能力的大幅提升也不可或缺。2007年，牛津大学人类未来研究所举办的全脑仿真研讨会提出了一个“路线图”，详细讨论了仿真需要的技术和恰当的层次。从理论上说，大脑仿真可以在多个层级或尺度上执行，包括量子、分子、突触、神经元或神经元群。根据大脑使用的确切机制，仿真可能必须在差不多的精细层级上执行。仿真的层级对仿真所需的计算资源有着重要影响。量子或分子层级的仿真可能会永远让人无法企及（除非量子计算取得极大进步）。从计算角度看，更高层级的仿真效率更高，但也可能会忽略一些对大脑行为至关重要的影响。

该路线图还包括一些里程碑，可用于衡量我们距离全脑仿真还有多远。它的结论（多少有点儿乐观）是，要在21世纪中叶实现全脑仿真，计算能力不足可能是我们面临的最大障碍。不过，这个结论建立在这样一个假设的基础之上：到21世纪中叶，所有的支持技术都已经足够先进。但是，同意这个假设的人寥寥无几。

这个路线图提出的一个重要观点是，全脑仿真并不一定需要创建全新的技术，对现有技术进行尺度上推（和大幅改进）也可以满足需要。毕竟，我们已经拥有可以对小块脑组织实施逆向工程的技术，基于这些模型构建的模拟器可以完成较小尺度的模拟。我们现在需要做的就是改进这些技术，先将它们应用于简单的无脊椎动物（比如秀丽隐杆线虫），然后应用于体型更大的动物（包括复杂的哺乳动物），直到最终仿真完整的人脑。这些发展都不需要全新的技术，但它们都需要现有技术取得巨大进步。

思维上传的第二个方法是，逐步置换工作大脑中的神经元。与大规模复制的方法相比，这个方法所需的技术更加复杂，距离我们也更加遥远。其中一项是纳米技术，这项技术目前尚处于起步阶段，但在未来可能会发挥重要作用。

1986年，埃里克·德雷克斯勒在他的有影响力的著作《创造的引擎》（*Engines of Creation*）中普及了一个观点：我们可以大量组装分子大小的机器，并利用它们完成现在不可能完成的各种任务。这个想法可能源于1959年理查德·费曼在美国物理学会年会上发表的演讲《在底部还有很大空间》（*There's Plenty of Room at the Bottom*）。在那次演讲中，费曼提出了直接利用单个原子创造出具有全新特征的材料和机器的可能性。

上述先驱者的想法被科学界朝多个方向加以发展。一些研究人员提出构建纳米设备的技术，我们可以将这些设备连接至神经系统，获取关于神经系统的详细信息。接下来，纳米机器群可用于检测神经脉冲，并将信息发送至计算机。最终，纳米机器可用于识别特定神经元，并利用其他纳米机器构成的阵列逐步置换这些神经元。于是，在大脑保持运行的同时，纳米机器可以依据局部的环境条件调整它们的行为。纳米机器只能依靠电磁、声音、化学或生物能源来驱动，有人已经提出了相关方案。

近几十年来，纳米技术取得了诸多可见的发展。很多直径只有几纳米的纳米颗粒已经在实验室中被创造出来，比如，直径只有1纳米的碳纳米管，还有纳米转子、纳米悬臂梁等。可以对单个血细胞进行分类和操控的芯片也被开发出来，即将投入使用，这使识别和清除血液中的癌细胞成为可能。但是，目前并不具备设计自主纳米机器人群（可以执行插入活体大脑或置换特定神经元所需的复杂任务）的技术，可能还要再花几十年甚至几百年的时间。

在尼尔·斯蒂芬森于1995年出版的小说《钻石年代》中，人们利用纳米机器人群将人类大脑相互连接成一个集体思维。我们与他描述的那个世界相距甚远，但没有显见的理由表明，这种技术永远不可能实现。

阿诺德·祖波夫在《一个大脑的故事》（*The Story of a Brain*）中，提出了一个与这个问题相关而且非常有趣的思想实验。在

这个实验中，一个人因病去世，但他的大脑没有受到影响。一个研究团队将他的大脑放到一个防腐槽里，然后模仿他活着的时候给他的大脑施加同样的刺激。随着实验的进行，更多的团队加入进来，这个大脑被多次分割，变得越来越小，最终成了一个个神经元。所有这些神经元都由高速通信设备连接，以确保每个神经元都能受到适当的刺激。最后，原来的那个大脑不见了，只剩下一个巨大的电子设备网络，为这些神经元提供输入。这些神经元与原来的大脑没有任何关系，而且它们本身是可模拟的。但是，读者可能会得出结论：尽管大脑被放进了防腐槽里，但那个人原本的个性未受影响，他的人生经历未被改写，他的情感也未被更改。

无论采用哪种方法，如果有一天我们可以在计算机上仿真人脑，我们将不得不面对若干难度极大的问题。我们之所以为人，是因为信息处理让我们能够感知世界，去思考、感受和表达我们的情感。支持这种信息处理能力的物理系统可以建立在生物或电子数字设备的基础上，所以不难想象，大脑的一个小部位——也许是视神经——可以在不影响大脑主人的主观体验的情况下被置换。但难以想象的是，在不影响大脑主人的主观体验的情况下，整个大脑可以被一台运行非常详细且精确的仿真程序的数字计算机置换。

终有一天，超级计算机可以运行一个完整的人脑仿真程序。为使仿真更加真实，需要给感觉系统（包括视觉通道、听觉神经，也许还包括传递其他感觉的神经通道）提供适当的输入。同样地，负责激活肌肉的大脑输出也应该被转化成适当的动作。如果没有这些组件，就会缺少真实的刺激（没有这些刺激，大脑无法正常工作），也就不可能精准地实现全脑仿真。

格雷格·伊根在2010年出版的小说《赞蒂奇》（*Zendegi*）中，设想了一项在不久的未来可能会出现的技术。身患绝症的马丁利用磁共振成像扫描了他的大脑行为，并用它们创造了一个合成智能体，期望这个智能体能通过游戏平台跟他的儿子交互。最终，这个计划失败

了，因为这个合成智能体——虚拟马丁——没有太多马丁的个性，以致无法在虚拟的世界中取代他做他儿子的父亲。基于马丁的大脑行为片段创造出来的虚拟马丁与人类相去甚远，这使研究人员认识到，目前还不可能将一个人的思维上传至虚拟现实。尽管结论是否定的，但这个想法引出了一些有趣的问题：我们什么时候才能确定复制的行为足够好？如何测试一个人的虚拟副本是否足以取代他本人，即便是在一些有限的领域里？如果副本不够好，你该如何处理它们？

可以说，如果真能做到全脑仿真，那么没有人可以理解工作大脑的完全成熟的仿真系统的所有细节。我们在做出任何一个关乎此类项目前景的决定时都必须考虑到，无论信息处理是在仿真系统中完成的，还是在活体大脑中完成的，都会造成同样的道德和哲学后果。

我并没有低估这类项目面临的技术挑战。我们谈论的不是模拟大脑的一个小的子系统，而是仿真完整的大脑，包括它承载的记忆、体验、反应和情绪等所有东西。我们距离实现这样的壮举还很遥远，在有生之年几乎不可能看到它实现。但是，这并不意味着这个目标永远不可能实现。事实上，我相信它总有一天会成真。

许多年甚至几个世纪之后，这种技术也许会出现。考虑到这种可能性的存在，我们必须清楚它的哲学、社会和经济后果。如果脑仿真器真的可以制造出来，它们很快就会大量出现，这是技术的本质。

开发全脑仿真所需的任何技术，都有很长的一段路要走。但有若干理由让我相信，我们必定能攻克这一难题。理由之一就是直觉，我认为这一领域的技术将以越来越快的速度持续发展。技术的指数性质往往会让我们感到惊讶，技术一旦起步，几乎总会产生出人意料的结果。现在用于理解大脑的许多技术同样如此，它们中有很多都是在过去几十年里发展起来的，虽然仅凭其中任何一项技术可能都无法生成重建大脑需要的所有信息，但技术的某种组合有可能发挥足够强大的作用。

我相信我们终有一天可以进行全脑仿真的其他原因，与这类技术的巨大经济价值有关。有些经济学家认为，世界GDP现在每15年左右增长一倍，在不久的将来每几周可能就会增长一倍。这些收益中的很大一部分都是由各种形式的数字思维创造的。

精确仿真工作大脑的能力将使我们对世界和社会的认知发生颠覆性变化，以至于我们目前面临的所有其他挑战在某种程度上都变为次要难题。因此，我们有理由相信，越来越多的资源将被用于发展这类技术，从而大大增加我们最终克服所有现存障碍的可能性。

意识之谜

如果复杂的通用人工思维出现，我们就不得不直面它们是否拥有意识的重要问题。

我们似乎都很清楚意识的含义。当我们在清晨醒来时就会恢复意识，并且在醒着的时候一直处于有意识的状态，直到晚上睡着以后再次失去意识。除了睡觉期间，有一种不间断的意识流把现在的你和多年前的你，甚至是刚出生时的你联系在一起，尽管大多数人的真实记忆都无法回溯到那么远的时刻。当被要求描述意识如何工作时，大多数人会分成两个主要阵营：二元论者和一元论者。

二元论者认为我们身处两个世界：一是物质世界，通过物理定律，我们对物质世界的研究和理解都较为充分；二是非物质世界，即我们的自我存在的地方。我们的本质（或者说灵魂）存在于非物质世界中，通过某种尚无法解释的机制与我们的肉体互动并控制着肉体。大多数宗教，包括基督教、伊斯兰教和印度教，都建立在二元论的基础之上。佛教是一个例外，因为它认为人没有永恒的自我，但不同的佛教流派，包括藏传佛教，对人死以后自我究竟有哪些特征继续存在持有不同的看法。二元论并不是宗教独有的，新纪元运动也涉及思维具有非物质力量的理论。

最著名的二元论是由勒内·笛卡儿提出的。1641年，他在《第一哲学沉思集》中提出了一个观点（现在被称为笛卡儿二元论）：心灵和大脑是两个不同的东西，心灵没有物理实体，而大脑控制的身体却是物理实体，并遵循物理定律。笛卡儿认为心灵和身体通过松果腺相互作用。尽管笛卡儿对现代哲学产生了深刻影响，但没有任何证据表

明承载灵魂的非物理实体存在，也没有证据表明灵魂是通过小小的松果腺与身体相互作用的。

笛卡儿二元论是对托马斯·霍布斯的唯物论做出的回应，霍布斯认为人类的所有体验都来自身体内部的生物过程。他的观点在19世纪的科学界颇受欢迎，现在仍很普遍。事实上，今天很少有科学家认为自己是二元论者。如果你从未考虑过这个问题，可以通过几个有启发性的思想实验来判定你是二元论者还是一元论者。

例如，二元论者认为僵尸至少在概念上是存在的，除了没有意识以外，僵尸的行为与正常人无异。如果你相信僵尸可以创造出来（比如，利用复制死人并将原子逐个组装成人类的方法或其他方法），并且你认为尽管僵尸不是人，但僵尸各个方面的行为都与真实的人没有区别，你可能就是一个二元论者。这样的僵尸会说话、走路、移动，行为举止也像一个有思维和灵魂的人；但它不是一个有意识的实体，而只是在模仿人的行为。

僵尸必须拥有情感、自我意识和自由意志，其行为举止才会与真实的人相似。它必须能根据它自己的心理状态和经历做决定，而且这些决定必须与真实的人做出的决定无异。雷蒙德·斯穆里安的小说《一个不幸的二元论者》阐明了一个因为信奉二元论而导致的悖论。在这个故事中，一个二元论者想自杀，但又不想让他的朋友难过。他得知有一种神奇的新药既可以消灭灵魂，又可以让身体像以前一样正常行动。于是，他决定第二天早上服下这种药。在他不知道的情况下，一个知晓他愿望的朋友在夜里给他注射了这种药。第二天早上，没有灵魂的身体醒来后去了药店，（再次）服下这种药。最后他愤怒地说，这种药没有效果。

这促使我们得出结论：如果一个生物的行为与人类一模一样，表现出来的情感完全相同，决策方式也没有任何不同，那么这个生物肯定有意识，也有它自己的思维。否则，没有人真正拥有意识，意识只是一种错觉。

唐纳德·戴维森基于另一个思想实验提出了相反的观点。假设你去沼泽地远足时不幸被闪电击中身亡。与此同时，在你附近的沼泽地里出现了另一道闪电，它自发地对一群不同的分子进行重新排列，以一种完全偶然的方式重新创造出你的身体，而且与你死亡那一瞬间的身体一模一样。戴维森把这个存在称作沼泽人，它在结构上与你一样。可以想象，它的行为也和你一样，要是你没有死的话。戴维森认为，尽管旁观者注意不到，但你和它之间必定存在显著的差异。沼泽人似乎认识你的朋友，但它无法认出他们，因为它以前没有见过他们。所以，沼泽人只是一个僵尸。在我看来，这个推理是荒谬的，是意识现象的二元论观点的直接结果，和笛卡儿几百年前提出的观点没有什么不同。

相反，一元论者不相信存在二元现实。他们认为，明显荒谬的僵尸概念表明二元论是完全错误的。术语“一元论”最早是由克里斯蒂安·沃尔夫在18世纪提出的，目的是解决二元论中存在的难题。一元论要么认为万物的本质是精神（唯心主义），要么认为万物的本质是物质（唯物主义）。（在这里我不讨论唯心主义，因为它与当今科学处理和研究物理实在的方式存在显著分歧。）

无论是否公开承认过，今天的科学家本质上都是唯物主义者。他们认为，所有可观测的现象，无论是意识、自由意志还是自我的概念，都是由一种物理实在产生的。

即使你可能反对二元论，但你仍倾向于认为有一个内在自我，持续不断地从某些特定方面关注你周围发生的事情。比如，你正在开车，并做着驾驶需要的所有事情，但你的意识可能在别处，也许正想着即将到来的假期或者稍后要交付的货物。内在自我在我们醒着的每一刻都保持着持续的注意力流，并构建着不间断的意识流。事事皆如此，就好像我们每个人体内都有一个小实体，观察着我们的感觉带来的一切，但它只专注于这些输入的某些特定方面，也就是会变成意识流的那些方面。这种观点有一个问题：据我们所知，并不存在这样一

个观察输入，即可以根据它的自由意志做出决策，并根据这些决策采取行动的实体。我们不可能通过移除大脑中的某个部位，使人除了失去意识和自由意志之外，其他一切还能保持正常运转。大脑中根本没有这样的部位，甚至所有可检测到的活动模式都不能扮演这个角色。大脑中存在的活动模式都是对特定输入做出反应，并产生一系列有计划的行动。

本杰明·利贝特的著名实验表明，我们对有意识的行动决策的感知欺骗了我们。利贝特要求受试者自发且有意地用一只手腕做一个特定动作，与此同时他记录下这个动作发生的确切时间，即大脑活动模式发生变化的时间，以及受试者报告他们决定移动手腕的时间。令人惊讶的结果并不是行动决策先于行动本身（约200毫秒），而是运动皮质（皮质的一部分，负责控制运动）中的脑活动发生变化的时间比受试者报告的行动决策时间早大约350毫秒。其他研究团队以多种不同形式再现了这个实验结果，它表明人对有意识的决策的感知发生在大脑启动决策过程后几分之一秒的时间内。因此，有意识的决策并不是行动的触发因素，恰恰相反，决策似乎是行动的事后证明。虽然这个结果本身不能证明负责一切有意行动的有意识的决策中心不存在，但它确实可以证明意识在某些方面具有欺骗性。与该实验结果一致的观点是，将意识仅看作我们对自我的感知，以及我们的行动改变我们的生活环境的方式。

很多有影响力的思想家都将意识现象视为一种错觉，即我们的大脑试图理解它的不同部位分别感知或引发的一连串事件。丹尼特在《意识的解释》中提出的意识的“多重草稿”模型，正是基于这一思想。在这个模型中，任意时刻都有各种事件发生在大脑的不同部位。大脑本身不过是几百万年的进化塑造出的“一揽子半独立机构”。这些事件的影响会传播到大脑的其他部位，并在事后被整合成一个连续的故事，也就是我们感知到的意识流。在这个模型中，有意识的“我”与其说是做出决定和发布命令的自主智能体，不如说它是一个记者，利用多个独立的平行事件构建出一个连贯的故事。在某种程度

上，这个模型可能与僵尸存在的观点一致，因为在通常情况下，事后将一系列事件的描述整合到一起未必会使智能体的行为发生多大的改变。

这个模型符合意识可能是一个历史不太悠久的现象的观点，这个观点大约是在几千年前出现的。1976年，朱利安·杰恩斯在他的备受争议的著作《二分心智崩塌中的意识起源》中提出，在公元前1000年前，人类没有意识，主要基于直觉并按照他们听到的“内心的声音”行事，他们认为这个声音是神的指示。根据杰恩斯所谓的二分心智理论，原始人的大脑被明确地分为两部分：一部分负责表达想法，另一部分负责倾听和行动。二分心智的终结——可能是强大的社会压力和更高效地实现长期目标的需要所致——标志着内省活动和人类意识的开始。

我们可能永远不会知道人类意识究竟是何时以及如何出现的。但随着技术的进步，更客观地理解意识是什么将变得越来越重要。确定特定的行为是有意识的还是无意识的，会对我们关于很多事情（包括犯罪行为）的判断产生很大的影响。

意识问题与哲学的另一个核心难题——自由意志——紧密地交织在一起。我们是能通过做决定来确定行动路线从而控制自己命运的实体，还是只能遵循由物理定律和接收到的输入为我们预先确定的路径的复杂机器呢？如果不存在有意识的我从大脑之外的某个地方做出决定并控制接下来采取的行动，我们又怎么可能拥有自由意志呢？

自由意志面临的一大障碍是决定论，该理论认为，在任意时刻都确实存在一个物理意义上的可能未来。经典物理定律为初始状态已知的所有系统都规定了确切的未来进化路径。如果你知道系统的确切初始条件和输入，原则上就有可能十分确定地预测出它未来的进化路径。只要它们是纯粹的物理系统，就不会给自由意志系统留下任何空间。

量子物理学和混沌系统理论给系统未来的进化增加了一些不确定性和随机性，但仍然没有给自由意志留下空间。事实上，量子力学给这个问题增加了真正的随机性，并改变了决定论的定义。量子力学认为，存在许多可能的未来，尚未展开的确切未来在波函数坍缩的那一刻才能确定。在波函数坍缩之前，许多可能的未来都有可能发生。当波函数坍缩且特定粒子进入某种状态时，就选定了一个未来。根据该理论的解释，要么一种可能的未来被随机选中，要么（在量子力学的多世界诠释中）这些可能的未来在平行宇宙中同时展开。无论是哪种情况，灵魂或意识都无法利用这种真正的随机性去控制未来，除非它能利用这种随机性去选择它想要的特殊未来。

尽管存在这些显而易见的困难，但一些人认为大脑或许可以利用量子不确定性，赋予人类一种独特的能力去对抗决定论，并施行他们自己的自由意志。在《皇帝新脑》一书中，罗杰·彭罗斯提出人类的意识不是一种算法，因此大脑不是图灵机的等价物。彭罗斯猜测量子力学在人类意识中扮演着至关重要的角色，某种在量子物理学和经典物理学的交界处发挥作用的未知机制使大脑比所有图灵机都强大，它也为意识创造了容身之所，并摆脱了决定论与自由意志相对立的困境。尽管这一论点可能很巧妙，但没有证据证明存在这样一种机制。事实上，相信这一解释的人归根结底都是二元论者。

也有一些人认为混沌理论是自由意志可以钻的漏洞。在混沌系统中，系统的未来进化完全取决于精确的初始条件。初始条件即使有非常微小的差异，也会导致系统的未来进化路径截然不同。在实践中，由于初始条件永远无法达到所需的精度，所以它无法预测未来。我们有充分的理由认为，大脑从很多方面看都是一个混沌系统。视觉输入的微小差异可能会产生完全不同的感知和未来行为。大脑的这个特征已经被很多人用作全脑仿真不可能实现的理由，甚至被视为自由意志存在的一种可能机制。但是，这些观点都没有引起人们的重视。无论是真正的随机性还是表面上的随机性，都不可能是自由意志的源泉，也不是大脑仿真的任何过程的障碍。的确，仿真大脑的行为或许会迅

速偏离真实大脑的行为，这可能是量子现象作用的结果，也可能是略有不同的初始条件造成的。但这两种行为都是合理的，并且都对应于真实大脑的可能行为。在实践中，无论是从主观角度还是从客观角度，我们都无法区分这两种行为中哪一种是真实的。

由于量子物理学和混沌理论没有为自由意志提供可乘之机，所以我们每个人要么是二元论者，要么必须找到某种方法来调和自由意志和决定论之间的矛盾，后者是一项困难但并非不可能完成的任务。丹尼特在他的发人深省的著作《自由的进化》中就曾尝试调和自由意志与决定论的矛盾，他指出，尽管在严格的物理学意义上，我们的未来是预先决定的，但我们仍然可以自行决定哪些东西比较重要，因为进化赋予了我们自己做决定的内禀能力，除了要遵循自然法则以外，不受其他任何强迫行为的影响。对丹尼特来说，自由意志就是做决定的自由，而不是背离物理定律的不可能实现的自由。

有一个思想实验可以帮助你判断你本质上是不是一个二元论者，即使你不认为自己是二元论者。德里克·帕菲特在《理与人》一书中请读者想象一种远程传送器，这种机器可以瞬间且无痛苦地对你进行破坏性的三维扫描，然后将信息发送给一个先进的远程3D（三维）汇编程序。汇编程序能够利用相同元素的完全不同的原子，纤毫不差地把你重新创造出来。有了这样的机器，你就能去到传送器可以送达的宇宙中的任何地方，但你是否愿意进行这样的光速旅行呢？如果你是一个真正的一元论者，并且认为物理实在可以解释一切，那么你只会把这样一台机器看作一种非常先进和便利的交通工具，就像查尔斯·斯特罗斯在他2006年出版的小说《玻璃屋》中描述的组装门一样。而如果你倾向于二元论，那么在使用这种旅行工具时你会三思而行，因为在另一端出现的你可能会失去一个重要的特征——你的灵魂。《玻璃屋》中的组装门根据真实物体的详细描述或3D扫描图像，利用纳米技术，逐个原子地重新创造有生命的和无生命的物体。组装门是一种通用汇编程序，可以按照给定的规约，用原子构建任何东西。组装门也创造出一种可能性，即一个人拥有多个工作副本，这是一个更加令

人困惑的现象。当然，远程传送器和组装门并不存在，短期内也几乎不可能出现，但它们引出的问题将会出现在数字领域，而且比人们预想得更快。

关于意识，这些会带给我们什么启示呢？除非你是一个坚定的二元论者，或者你相信意识具有非算法性质（据我们所知，这两种立场都很难自圆其说），否则你就只能承认僵尸不存在。如果某个实体表现得像有意识的存在，那么它肯定有某种意识流。在这种情况下，意识流是一个足够复杂系统的涌现性质。

从概念上讲，意识可能有多种不同的形式。托马斯·内格尔在他的著名论文《做一只蝙蝠是什么感觉？》中，请读者想象如何从蝙蝠的视角看世界。当然，想象你是一只蝙蝠，一边在黑暗中飞行，一边通过雷达感知周围的环境，这样做没有任何意义，因为一只受过教育和会讲英语的蝙蝠肯定不是一只普通的蝙蝠。想象你不用眼睛，而是通过声呐来感知世界，这样做也没有任何意义，因为我们不知道蝙蝠是如何感知世界的，也不知道蝙蝠的头脑是如何描绘这个世界的。事实上，我们甚至不知道在其他人的脑海中这个世界是什么样子，尽管我们可以假设，别人头脑中的世界可能和我们头脑中的世界差不多。

内格尔问题的重要意义在于，是否有跟做一只杯子不一样的体验，比如，做一只蝙蝠。没人会真的相信人变成杯子这种事，但关于蝙蝠问题的讨论并未尘埃落定。和其他人一样，内格尔也断定关于意识到底是什么的问题永远找不到答案，因为它建立在对完全主观经验的客观理解的基础之上，这本身就是一种自相矛盾的说法。要想知道成为蝙蝠是什么感觉，你就必须成为一只蝙蝠，而蝙蝠是无法向除蝙蝠之外的生物描述它们的体验的。要解决这个问题，蝙蝠就必须用人类的语言和情感描述它自己的体验。然而，这样的蝙蝠绝不是一只真正的蝙蝠。

我们似乎又回到了起点，而不是离理解意识是什么更近一步。因此，如果我们不知道某些系统用来处理信息的确切机制，就必须求助

于某种客观的黑盒测试，比如图灵测试。相反，如果我们知道某个系统与另一个系统执行完全相同的计算（即使它的计算基底不同），那么除非我们是坚定的二元论者，否则就必须承认这样的系统是有意识和自由意志的。

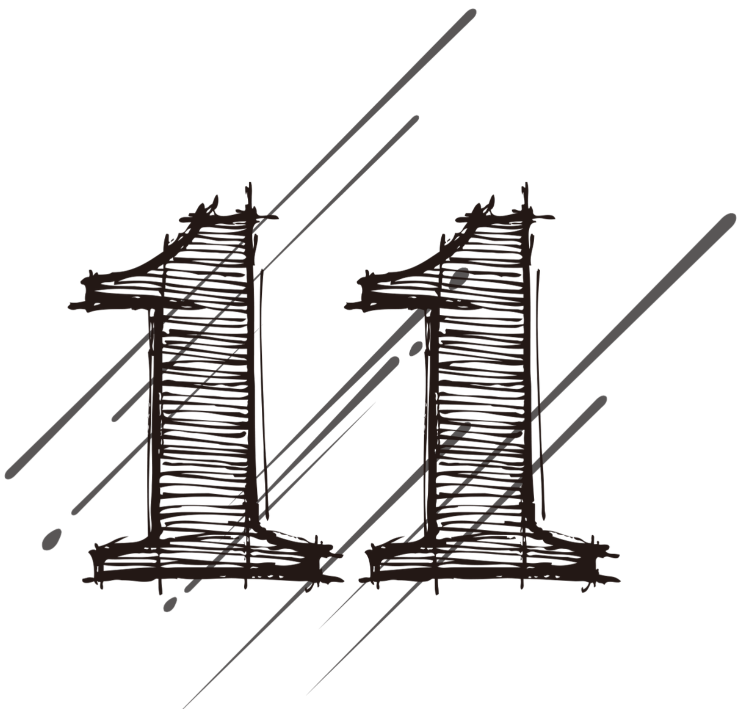
因此，受大脑行为启发或从活体大脑复制而来的智能系统，根据其定义，几乎就可以确定它是有意识的。当然，我指的是成熟的智能系统，它的结构是按照与人脑相同的原理进化而来的，或者是从功能健全的人脑一点儿一点儿复制而来的。

根据完全不同的原则以自上而下的方式设计的系统是否会引发一系列新问题？对于像谷歌这样的搜索引擎或者像沃森这样的专家系统，我们会不会认为它们拥有意识呢？为了让这样的系统更人类化，它们的设计者是否会为其提供更智能的界面，让它们能记住过去的交互并拥有自己的动机和目标呢？未来某一天，现有的某些系统可能会具备这种能力，但我们显然不会认为它们拥有意识。这是因为它们缺少某些重要组件（或许是一些小组件），还是因为我们持有以人类为中心的立场，所以拒绝承认它们拥有意识呢？我认为这两个原因兼而有之，除此之外，它们的设计者不愿意提出这个问题，以免激起人们对超人类智能的恐慌情绪，这对商业经营不利。但是，当真正的智能系统普遍存在时，人们迟早要面临这些深刻的哲学问题。

下一章将讨论这些棘手的问题。由于技术的发展会带来千奇百怪的挑战和问题，所以下一章的内容可能会让你觉得完全脱离了现实。如果这一章的观点让你觉得无法容忍，那么你最好不要往下读了，因为与后面两章相比，本章的内容真的不足为奇。

第11章

挑战和希望



除非你是坚定的二元论者，否则你肯定不会认为数字思维永远不会出现。事实上，尽管我们现在尚未掌握创造数字思维所需的技术，但也很难相信这些技术永远不会出现。我不想去预测这些技术何时将被开发出来，但原则上我们没有理由怀疑某种先进的数字思维将在21世纪成为现实。

公民权利

如果一个系统被视为数字思维，我们就必须考虑公民权利这个棘手的问题。现有立法承认每个人都具有一种特殊身份——人的身份，对我们来说这是理所当然的事。今天，人格以及随之而来的权利和责任，已经被赋予了几乎每一个社会中的每一个人。但是，人格的概念不是普适的。此外，不论种族、性别或地位，普遍承认每个人都具有人格权，这在历史上是较为晚近的事情。在奴隶制度被废除之前，奴隶不被视为人，废除奴隶制度是一个漫长而缓慢的过程，历时数百年。在美国，赋予各州女性投票权的宪法第十九修正案直到1920年才获准通过。

人格的概念因国家法律体系的不同而有差异。在美国和其他许多国家，人格得到法律的承认，因为人拥有权利和责任。人是权利和责任的法律主体或实体。人被称为“自然人”，但人格也可以被赋予公司、协会和集团。

只要赋予数字思维权利和责任，就会立刻引发关于其人格和身份的讨论，这种假设看起来有一定道理。数字智能体——代表个人和公司做出重大决策的软件程序——已经存在，许多影响人类生活的事件都是它们负责的。基于一系列复杂规则和算法的自动交易智能体，是最显眼的数字智能体，它们对经济产生了巨大的影响，并被视为经济危机和其他棘手事件的罪魁祸首。但直到现在，当数字智能体造成不良后果时，承担责任的并不是它们，而是拥有它们的个人或公司。造成这个结果的原因很简单：截至目前，人格还未被授予任何纯粹的数字实体。

但是，随着数字智能体及其行为变得越来越复杂，承担责任者自然会从它的“所有者”变更为它本身。事实上，如果数字思维被赋予完整的人格，那么智能体所有权的概念终有一天会成为争论的焦点。整个过程将花费很长时间，其间也会引发很多讨论，但最终社会不得不承认数字思维是自主的“人”，拥有它们自己的动机、目标、权利和责任。如果数字思维是通过思维上传的方式创造出来的，那么数字思维是否具有人格的问题将无法逃避。思维上传技术在21世纪末之前有可能实现，所以社会必须解决这些数字思维的人格和身份问题，为它们提供一些公民权利和责任。我将把这些实体简单地称作“数字人”，并用它来指代被赋予某种人格的所有数字思维。为简便起见，我使用代词“它”来指代数字人，但这并不意味着这些人的人格弱于真人，也不意味着它们没有性别。无论它们认为自己是男性还是女性，抑或两者都不是，这取决于许多因素，我在这里不做阐述，因为涉及的变量太多了。结合具体情境，我也可以用“他”、“她”或“它”指代数字人，但这会让我们觉得尴尬。

最核心的问题可能与数字人的权利和责任有关。假设一个程序凭借其先进的语音和视觉界面，系统、永久、反复地让评委相信它的行为与人类一样，从而通过了全面的图灵测试。要做到这一点，程序必须展现出人类在类似情境下展现出的推理、记忆和学习能力，甚至是情感。这个程序——一个数字人——的大部分知识都是通过技术手段获得的，有助于改进它的行为，并随着时间的推移变得与人类的行为越来越相似。如果一个程序不具有某种形式的意识，那么我们很难想象它可以通过全面的图灵测试，但现在我们可以假设它只是一个非常像人类的僵尸。

数字人或许会在社会中扮演许多重要的角色。至少在原则上，它们可以产出由人类创造的任何智力成果，也可以驾车、开飞机、撰写新闻或为客户提供服务。事实上，很多诸如此类的任务现在通常都是由程序完成的。我们今天看到的许多新闻都是为美联社等新闻机构工作的机器人记者撰写的，自动驾驶汽车每天都会出现在新闻中，飞机

自动驾驶技术已经使用了多年。包括Clara Labs（克拉拉实验室）在内的多家公司，都在网上提供可帮助客户安排会议的虚拟助理。几乎每一款新手机的操作系统中都包含像Siri（苹果智能语音助手）或Cortana（微软小娜）这样的虚拟助理，任何人都可以使用，但虚拟助手仍有很多缺点，也有很多请求是它们无法处理的。

即使没有独立于计算机的物理存在，数字人也可以写信、买卖股票、打电话、操控银行账户，以及进行许多人类特有的但不需要物理干预的其他活动。因此，发展数字人并利用它们去促进经济活动，具有强大的经济动机。在许多方面，与自然人相比，数字人并不处于劣势地位。它们可以像自然人一样开展媒体宣传活动，召开媒体发布会，上电视，并影响公众舆论。

数字人应该拥有哪些权利呢？它们应该被赋予人格以及随之而来的所有权利吗？它们应该像奴隶一样归个人或公司所有，还是应该成为它们自己命运的主人？它们可以拥有金钱和财产吗？它们有投票权吗？如果更智能、更灵活的新模型或新版本出现，导致它们不再有利用价值，它们能否被终止（从存储器中删除）？

至于数字人的责任问题，同样不容易回答。假设一个数字人触犯了法律。它应该承担罪责，并被送进监狱吗（暂且不讨论这里的监狱指的是什么）？或者，因为这些程序是程序员多年前编写的，使得数字人获得了知识和个性并导致了当下的违法行为，所以应该把这些罪责归咎于程序员吗？如果数字人意外或故意致人（无论是不是自然人）死亡，它应该被判处死刑吗？在我撰写本书时，美国国家运输安全委员会正在分析一辆特斯拉S型汽车在自动驾驶模式下与一辆白色卡车相撞的事故，由于它没有探测到这辆卡车进而及时刹车，结果导致它的人类乘客死亡。随着这些技术的不断普及，其他类似的案例可能也会接踵而至。

如果一个程序通过了全面的图灵测试，让所有人都相信它可以像人类一样思考，那么应该把大奖颁给创建它的程序员，还是这个程序

呢？如果把大奖颁给程序员，程序员可能会非常高兴，但程序可能不会同意这样做。毕竟，如果一个程序的人性化程度足以让它通过图灵测试，那么它是否也有权分享大奖呢？

尽管这些问题看起来稀奇古怪，但它们并不是对永远不会存在的假想实体的闲来无事的思考。数字思维迟早会存在，一旦它们出现在这个世界上，关于它们是不是数字人的讨论就不可避免。可以肯定的是，如果真正的数字人出现了，那么它们给社会带来的挑战将比我们今天关注的最具争议性的问题还要复杂。

原件、副本和复制品

数字世界和物理世界的一个主要区别是，数字实体只存在于计算机存储器中，易于复制、存储和检索。在物理世界中，人们非常重视原件与副本之间的区别。银行柜员会很乐意兑现一张签名支票，但如果他见到的是同一张支票的两份副本，甚至是两张不同的支票，而且第二张支票上的签名是第一张支票签名的扫描版，他就不太高兴了。在餐馆吃完晚饭后，用50美元的真币埋单和用50美元的假币（即使仿造水平很高）埋单，收银员对你的态度将截然不同。毕加索的画作和其赝品的价值存在着天壤之别。

这种状况与物理世界中对象的完美复制面临的困难相符。直到最近，物理对象的精确复制仍然很难甚至不可能做到，而且社会的运行依赖于一个假设，即物理对象的完美副本很难得到。伪造是有可能的，但通常可以检测出来，而且几乎都是非法的。处理货币、签名、指纹和所有其他个人身份识别方法的技术都基于一个事实：物理对象的精确副本很难创建。要精确复制人体或人体的某些部分，比如手指或视网膜，难度尤其大，因此利用生物测定方法核实自然人的身份，可信度非常高。事实上，手写签名之所以是最常见的身份识别方法之一，其根本原因是复现人手完成物理签名的动作序列的难度极大。

但是，基于物理对象拷贝难度大的认证和识别机制面临的局面正在迅速发生变化，原因主要有两个。第一个原因是，精确复制不像以前那么困难了；第二个也是更重要的原因是，在数字领域复制很容易做到。

目前正在开发的技术将使我们在不久的将来可以对物理对象进行高精度扫描，并使用所谓的通用汇编程序（可能还需要使用纳米技

术），去组装任意数量的副本。未来某一天，纳米技术或许能让我们根据以数字形式存储的规约，逐个原子地组装对象。这些技术在付诸实践时会给我们的物理世界带来一些有趣的挑战，但这些对我们当前的讨论而言不太重要，因为在通用汇编程序出现之前，与之类似并且复杂得多的问题可能早就在数字世界中出现了。

在数字世界中，精确的副本很容易获得。我们都熟悉今天的数字技术带来的挑战。软件易于复制，一切以数字格式存储的媒体（比如书籍、音乐、视频和照片）皆如此。所有基于创建、拷贝和分发媒体内容的行业，包括报纸、图书、电影和音乐产业，都需要适应数字世界，在这个世界中，数字对象的精确副本易于制作、获取和广泛分发。有些商业模式可以适应，有些则无法适应，但这些行业都在发生深刻的变化，仅仅10年前还存在的东西，几乎不可能一成不变地再存在10年。报社、出版社、电影公司、词曲作者和歌手，都必须应对人们在生产和消费内容的方式上发生的深刻变化。

在数字世界里，原件和副本没有任何区别，数字实体的本质使得某个实例到底是原件还是副本变得无关紧要。在某些情况下（比如，当我们拍摄一张数码照片并将其存储在相机卡里时），我们可以确定构成数字实体的位序列最初是在哪里汇编的。但在有些情况下，我们很难甚至不可能确定某个数字实体最初出现在哪里。例如，当我们通过电话发送语音信息时，信息会通过多个分散的数据包在空中传播，然后被记录在云中，后期还可以复制。但在所有这些情况下，“最初”的位序列与它形成的副本都无法区分。存储在你计算机中的数码照片副本与相机卡第一次记录的那个版本没有优劣之分，你发送给朋友的所有副本不仅品质相同，而且从某种意义上说，它们和你存储在相机卡中的版本不存在原件和副本之分。在考虑这些问题时，我忽略了数字实体中的误差累积问题，因为这些误差非常少见，也很容易纠正。

复制数字人的可能性引出了大量复杂的哲学问题。事实上，人们可能对数字思维的最终可行性和存在性持有发自内心的怀疑态度。但是，我们完全可以放心地假设，如果数字思维真的出现了，备份和存储它们以备将来检索就是切实可行的，而且总的来说，可以任意多次地精确拷贝它们。基底的计算支持与创建及开发数字思维的精确技术无关，它可能取决于标准的数字存储，数字存储的本质允许数字思维被精确拷贝任意多次。

二元论者总会辩称，尽管生成思维的系统的数字支持可以拷贝，但思维的本质却无法拷贝。相信量子现象在意识出现的过程中起着重要作用的人会感到些许宽慰，因为我们无法在不干扰原初系统的前提下复制系统的量子组态，而数字实体可以在不扰乱原初系统组态的条件下被复制。但是，并没有证据证明这些观点是正确的。如果数字思维真的出现了，它们可能会易于拷贝和复制。我们很难理解这种可能性到底意味着什么，因为在物理世界中不可能出现的许多情景，在数字世界中不仅变得有可能，而且寻常可见。

接下来，我们考虑直接拷贝数字人的可能性。现在想象在21世纪的某个时候或22世纪，一种使非破坏性的思维上传成为可能的技术被开发出来，而且上传的思维获得了数字人的身份。

我们必须先处理如何对待这个数字人和作为上传模板的生物人的问题。在刚刚完成复制时，这两个人（数字人和生物人）的行为是相同的，或者至少非常相似。副本将生活在某种虚拟环境中，它希望可以访问一些资源，以便让自己生存下去。也许它想要钱去购买运行仿真程序的计算机时间，或者购买其他物理资源和数字资源。生物人（想必是）同意被扫描和复制到计算机的存储器中，他可能会以一种同情的眼光看待数字人的这种欲望，甚至可能愿意将自己的资源分享给数字人。到复制完成的那一刻，两个人的记忆、情感和经历完全相同，但此后他们的记忆、情感和经历则会开始彼此偏离。差别将缓慢产生，并随着时间的推移不断积累。我们可以想象，在这种情况下，

复制可能类似于自我离婚，在自己和自己的副本之间分割财产；复制也可能类似于结婚，财产由自己和自己的副本共同拥有。

也许只有在被上传者的两个未来副本订立合适的且有约束力的合同的前提下，非破坏性的思维上传才有可能是合法的。当然，在启动上传操作时这两个副本还不存在，因此订立合同的做法有点儿难以实现。不过，这个问题并不严重，因为两个副本都与原件没有任何区别。上传操作刚完成时，两个副本拥有共同的过去和不同的未来。因此，可以在上传之前拟定并登记一份合同，以便约束签署合同的人的两个未来版本。自然人在上传操作启动之前签署的这种合同将强制执行未来的资产分配，并且应该对复制之后的两个（相同的）人都具有约束力，因为他们过去的自我同意并签署了这份合同。我们很难预测这个机制是否有效，但它似乎是确保这两个实体不会因为权利和资产纠纷而对簿公堂的最合乎逻辑的方法。有人也许认为这个方法无效，因为这两个未来的人（或者其中之一）可能会认为这份合同不可接受，但我们可以建立法律机制，把这种程序纳入合适的框架，毕竟双方都同意这样做。

另外，即使思维上传成为现实，它也有可能不得不使用某种破坏性技术。由于扫描技术的限制，要想获得可用于仿真的详细大脑副本，原初的大脑可能就难以摆脱被破坏的命运。这种情况不如前一种情况复杂，因为它不是真正意义上的复制，资产归属权的问题也应该更易于处理。继承者可能会声称它们的“前任”已经死亡，但如果排除这种可能性，那么被上传的数字人应该可以保留（或者说继承）原初的生物人的权利和资产。在撰写本书时，我假定数字人的法定权利和责任在这项技术变得可行之前就已经确立了。

一旦出现了数字思维，无论它是通过上传还是其他方法获得的，数字人的复制就会变得轻而易举。数字人的新副本的复制和实例化，是一个非常简单的过程，即使是现在的技术也有可能做到，但前提是数字人有一个计算机可执行的副本。这个过程面临的挑战与思维上传

面临的挑战类似，但在这种情况下，假定原初的数字人在这个过程中被破坏是没有意义的。同样地，在复制发生之前，可能需要签署一份对数字人的两个实例都具有约束力的合同。复制数字思维和上传生物思维的区别在于，前者存在全对称性，因为数字思维的两个副本都不能宣称自己更接近原件。而在思维上传的情况下，生物人是拷贝的基础，他或许有理由宣称自己是原件，所以应该有权决定资产的分配，甚至终止拷贝。

即使不上传思维，而是考虑思维设计的其他可能性，问题也不会变得更简单。如果数字人拥有权利，复制过程的操控肯定会非常棘手。最根本的难题源于一个关键的区别：生物人从生到死都在一条线上，不会出现分支，所以从出生那一刻起到死亡那一刻止，人格的定义都指向同一个人。虽然人体内的原子不断地被新原子替代，人的大部分细胞每隔几年就会被置换，但一个人的人格在本质上是不变的。毫无疑问，至少对社会而言，他与几年前的他是同一个人。罪犯即使在自我辩护时声称自己现在是另外一个人，也不能轻易逃脱罪责。但是，如果我们考虑的是数字思维，情况就会有所不同，因为在这种情况下，时间流变得不那么容易预测了。事实上，在有限却重要的意义上，时间有可能前进，也有可能倒退。

时间旅行、备份和回退

很多小说、电影和科幻故事都是以时间旅行的可能性为主题创作的。尽管有些人相信未来某一天某种机制或许会让时间旅行成真，但除了相对论效应，我们进行时间旅行的唯一方式就是以每秒钟一秒的正常速度走向未来，没有任何办法可以改变这个事实。也就是说，在我们知道和经历的这个世界里，我们随时间前行的速度只能和其他人一样，也和我们周围的物理世界的其余部分一样。

相对论预测，当一个物体相对于另一个物体运动时，两个物体的时间都会变慢；当一个物体相对于另一个物体以极高的速度运动时，这种效应会更强。时间延缓已经被观察和测量了很多次，事实上，它会导致全球定位卫星时钟和地球时钟的速率不同，只有校正这一差异才能使全球定位卫星系统正常工作。同样地，如果宇航员乘坐航天飞船以接近光速的速度飞行，那么他一生中可以看到外部世界几百年的光阴流逝。除非用这种方法实现时间旅行（这仍然是科学和技术无法企及的目标），否则我们就只能接受在我们日常生活的这个物理世界中，时间旅行根本无法实现的事实。但是，当我们用计算机仿真数字人的生活和思维，并考虑它们做时间旅行的可能性时，情况就不同了。

仿真数字思维不需要以物理时间的正常速度进行。我们可以根据可用计算能力的大小，来决定是以实时的速度、比实时快的速度还是比实时慢的速度，运行虚拟世界的仿真。为便于连接现实世界，我们有理由假定数字思维的仿真通常会以与实时相同的速度运行。这使它与物理世界的交互变得更容易，而且在很长一段时间内都非常重要。

对数字人（尤其是经历过物理世界的数字人）来说，在物理世界中观察到的时间流逝速度与它的主观时间大不相同，将会是一件奇怪的事。但是，时间流逝速度的差异可能是必要的，甚至是人们期望的，原因有很多。比如，我们可以想象，计算资源的缺乏需要计算机以实时速度的100倍运行仿真程序。在这种情况下，主观的一天就相当于物理世界中的一整个季节。在自然人的一个星期里，数字人将会看到两年时光的流逝。在和物理世界中的自然人进行对话和其他互动时，数字人必须放慢速度，双方才有可能顺畅地交流。还有一些现象也会让数字人感到别扭，在物理世界中每过一秒钟，数字人的主观时间就会前行100秒。这足以使它观察到外部世界里的孩子每天都在以肉眼可见的速度成长，朋友变老的速度也异乎寻常地快。

反过来可能也一样。如果有足够的计算机能力，我们就有可能以快于实时的速度仿真数字思维，从数字人的视角看，这实际上减缓了外部时间。在以100倍于实时的速度运行仿真程序时，数字人将看到自己生命中的两年过去了，而物理世界才刚刚过去了一周。同样，与外界自然人进行对话和其他互动时，数字人也会感到别扭和困难，每句话说完后都必须停顿很长时间，以便外面世界的生物人处理数字人说的话，因为数字人的说话速度要快得多。

正向的时间旅行还有可能采取其他更激进的形式。例如，数字人也许会决定一下子向前跳跃10年。只需要在指定的时间段停止仿真即可做到，也就是说只需轻轻按一下开关。对数字人来说，到10年后重启仿真程序时，根本没有时间流逝。关闭仿真程序后，从数字人的主观角度看，一眨眼的工夫外部世界已经过去了10年。

数字思维和数字世界的存在，也使我们有可能进行回到过去的时间旅行。我在前文中间接提到一个事实，即数字思维和其他任何数字实体一样，都可以随意拷贝。也就是说，在任意时间，我们都可以将数字思维的状态拷贝到某种存储介质中，以后（或许是多年之后）还可以恢复。这个简单的操作可以对数字人执行常规备份，并在将来的

任意时间将数字人恢复到与备份时完全相同的状态。通过这种方式，如果可以获得过去某个特定时间的备份，就有可能将任何人恢复到那时的状态。

任何熟悉计算机技术的人都知道，通过一种叫作增量备份的技术就可以对计算机执行常规且花费不多的备份。增量备份比普通备份更先进，它只存储自上一个备份之后发生的变化，这样一来，我们就能将计算机的状态恢复至任意时间点。如果用这样的技术确保数字人可被恢复至过去的任意时间点，那么从数字人的主观角度看，这个方法不仅可行，回到过去也会变得很容易。

这种可能性会带来许多有趣的结果。除非有特殊需要，否则数字人永远不会真正死亡，因为它随时可以回到过去，通过备份恢复到年轻状态。在查尔斯·斯特罗斯的小说《玻璃屋》中，致命决斗十分常见，而且得到了社会与文化的认可，这是因为决斗者随时可以通过备份恢复至决斗前的状态。可能发生的最糟糕的事情是，从最后一次备份到决斗的那段时间的记忆会丢失，这会造成些许不便，但并不严重。

现在我们认为只会出现在科幻小说里的其他可能性，也有可能成为现实。其中一个可能性是，数字人与年轻版本的自己或年长版本的自己进行对话。例如，如果一个数字人想让年轻版本的自己知道一些关于它的未来生活的有用信息，那么他只需通过备份恢复年轻版本的自己，然后与它进行一次私密谈话，其间或许还可以一起喝几杯啤酒。但此后，情况会开始变得混乱。如果年轻版本的数字人能够利用这些信息，它将抱持着这个见多识广的唯一信息来源提供的宝贵建议，继续生活下去。但这样一来，这个数字人将会有两个不同年龄的副本存在。或者，年长的副本觉得自己的人生历程充满遗憾，因此选择终止自己的生命，而让汲取了它的智慧的年轻副本成为新的数字人。实际上，这是数字人对它过去做出的决定感到后悔时可以采取的一种补救方法。例如，在一段失败的恋爱关系终止后，数字人可能会

决定结束自己的生命，然后恢复至以前的备份版本并重新开始生活。在电影《美丽心灵的永恒阳光》中，一个先进的程序可以选择性地删除不再需要的记忆。与电影中的情况不同，恢复以前的备份就相当于删除从备份完成那一刻起到数字人决定回到过去那一刻止的整个时间段的所有记忆（而不是选定的某几段记忆）。

这些奇怪的可能性带来的问题和挑战还不止这些，事实上，它们招致的麻烦几乎无穷无尽。假设一个历史清白的数字人犯了重罪，并被判处终止生命，那么能否允许它恢复至犯罪前留下的备份版本呢？如果可以，那么恢复之后的它是否应该对它的副本在它的主观未来（其他人眼中的过去）犯下的罪行负责呢？我们应该牢记，数字人出现后，严重罪行的性质将发生重大变化。在存有备份的情况下，只要被谋杀的人拥有近期的备份，数字人的谋杀行为就不太严重。事实上，谋杀可能会被视为一件稀松平常的事，后果轻微——相当于今日世界的一记耳光。自杀也将成为无效行为，因为数字人可以轻易地通过备份恢复至自杀前的版本。

在未来几十年里，随着创造数字人所需的技术得到发展，所有这些奇怪的新可能性和我未提及的许多可能性都将出现。不过，如果我们在真正面对这些问题之前就能理解它们，对我们将大有裨益。至少在某些特定方面，我们可能很快就会面对其中的一些问题。

生活在虚拟现实之中

说到这里，我们有必要深入探讨数字思维到底会如何与物理世界交互（如果这是真的）。到目前为止，我总是略过这种互动涉及的实际困难。我描述了一些与大脑仿真有关的技术问题，还描述了一些为数字思维的出现创造条件的其他计算智能体的设计相关的技术问题，但还没有阐述数字思维与外部世界交互的重要问题。对上传思维和神经形态系统来说，这是一个重要的问题，因为对大脑进行精确和逼真的仿真需要与物理世界进行细致的交互，否则就不可能让创建经验、记忆和情感的神经元之间建立适当的连接。

显然，外部世界的输入可以通过摄像机和传声器传送到视觉和听觉神经，从而刺激视觉和听觉。我们也可以通过激发其他感觉神经，为数字思维提供触觉、嗅觉和味觉等，使虚拟世界中的生活与物理世界中的生活类似。要让数字人受到一整套刺激，需要克服的工程难题可能和创造数字思维需要克服的难题一样复杂，但它们本质上属于技术挑战。众所周知，人类有5种基本感官，除此以外，人类还有其他感官，为我们提供来自物理世界的各种输入。不过，一个包含所有5种感官的精细仿真程序，就已经是一个大到值得讨论的挑战了。

另一方面，数字思维必须能通过控制产生语言和运动的肌肉来影响外部世界。控制这些肌肉的运动神经元，接收来自其他神经元（通常是脊髓中的神经元）的输入，然后向肌肉传递信号，从而产生运动。

人们正在积极研究有效模拟这一复杂过程所需的技术，因为这些技术也可以帮助残疾人控制假肢或用于其他许多目的。当新创建的数字思维接收到来自陌生信息源（比如数码相机、传声器等）的输入

时，大脑可塑性（大脑适应新刺激的能力）将会发挥作用，帮助大脑适应并学习如何生成正确的神经放电模式。大脑适应性已被广泛观测到并应用于实验环境，例如，让盲人的神经皮质受到视觉输入的直接刺激，让聋人的耳蜗受到听觉输入的直接刺激。但如何利用（模拟）运动神经元接收到的这些信息与外部世界互动，并将感觉装置移动到动作发生处，仍然是一个挑战。

更显见的可能性是，让计算机中的数字思维通过机器人与外部世界交互。先进的机器人技术就像电影《机械姬》中的人工智能机器人一样，能够可靠地模拟数字思维的预期动作和交互。机器人（无论是不是类人机器人）可以使数字人与物理世界交互。要让机器人拥有的一整套传感器和执行器去模仿人类拥有的传感器和执行器，需要耗费几十年的时间开发一系列技术，但随着机器人技术的进步，所有这些内在的技术难题都将迎刃而解。脑机接口的研发人员已经取得了多项重要成果，其中包括让四肢瘫痪的患者仅通过思维就能控制肢体的设备。未来某一天，这些技术将会帮助普通人控制计算机，或许还会取代现有的接口；数字思维也有可能利用这些技术来控制物理世界中的机器人。脑机接口研究人员开发的某些技术也可用于直接刺激脑区，创造出与触觉和其他感觉相同的感官体验。这些技术还可以为数字思维提供沉浸于物理世界的真实感。

另一种可能性不太显见，但却更灵活。数字人可以生活在一个完全模拟的数字世界中，在虚拟环境中模拟感官接收的输入和肌肉引起的动作。大多数读者可能都熟悉电脑游戏中常用的多种类型的虚拟现实。当今最先进的视频游戏已经达到了令人印象深刻的逼真水平，模拟足球或篮球的视频游戏在逼真程度上堪比真实世界的影像。当所有交互都发生在计算机屏幕上时，沉浸感是不完整的，但更具沉浸感的界面很快就会接近完全骗过人类感官的水平。

虚拟现实被许多人誉为未来人类与计算机交互的方式。例如，好几家公司提出制造虚拟现实眼镜，但到目前为止，它们还没有成为人

类与机器交互的主要方式。部分原因在于，虚拟现实设备难以准确地模拟视觉输入。如果运动和渲染的图像不一致，对此敏感的大脑就会引发不适感和定向障碍。不过，随着虚拟现实技术的发展，交互的质量也得到提升。Oculus VR等公司现在承诺，它们提供的虚拟现实体验与真实的视觉体验基本上没有区别。谷歌的创新性“纸盒”项目使任何拥有智能手机的人都可以随时感受虚拟现实，花不到20美元就可以买到一个纸盒，与智能手机应用程序结合使用，可体验到一种十分逼真的沉浸式虚拟现实环境。

到目前为止，这些沉浸式界面只能带来视觉和听觉的沉浸式体验。而当涉及身体时，触觉和运动等其他感觉的处理难度则要大得多，因为必须移动身体并将其置于沉浸式的特殊环境中。这需要体积庞大的执行器和传感器，所以很难实现。不过，对没有身体且只在计算机中运行的数字思维来说，这个问题更容易解决。在这种情况下，只要作用于合适的感觉神经，就可以产生运动（包括加速）的物理感觉，而且是十分逼真的沉浸感。

我认为，随着图像渲染技术的进步，以及我们对运动神经元和感觉神经元的运行机制的进一步理解，高度逼真版本的数字世界不仅有可能实现，而且最终会成为数字人与外部世界交互的选择机制。当然，这里的外部世界是一个非常特殊的世界，因为它完全是在计算机中模拟出来的。人潮涌动的城市，绿油油的田野，浪花朵朵的海滩，白雪覆盖的山脉，这些都可以被模拟到相当逼真的程度，以至于它们与现实世界难以区分。与其他数字思维的交互简单易行，因为这些思维都生活在同一个虚拟世界中。与真实物理世界的交互也可以进行，不过还需要其他设备。物理世界（或者部分物理世界）的状态，可以通过摄像机和其他传感器输入虚拟世界；运行虚拟世界模拟程序的计算机，可以对物理世界中的人和其他智能体进行适当的渲染。在外部的物理世界中与人类的真实物理交互仍然是一个限制因素，但这一问题可以通过使用机器人或这两个世界之间的某种同步机制来解决。

大量的数字思维生活在虚拟世界中并发生互动，这似乎是一个稀奇古怪和令人难以置信的概念，但这可能是为数字思维创造适当条件的最显见也是最划算的一个方法了。未来可能确实像格雷格·伊根在他的小说《置换城市》中描述的那样，用一块计算机芯片就可以仿真1000万人的身体和大脑。

在虚拟世界中，我们知道的物理定律未必适用，这虽然不是什么大不了的问题，但仍然值得我们注意。在一个纯粹的虚拟环境中，人们可以即刻从一个地方去到宇宙中的任何地方，只要目的地可以被模拟出来。在虚拟世界中，数字人不需要任何特殊设备就可以飞越高山，潜到海底，或者在月球表面行走。事实上，虚拟世界拥有无限可能，未来某一天到底会出现什么样的虚拟现实，这完全超出了我的想象力。对虚拟现实游戏的玩家来说，其中一些奇怪的可能性他们已经习以为常了。

在虚拟世界中，即使回到过去也是可以做到的。在计算机生成的虚拟现实内部，肯定可以从任意时间点重新运行模拟程序。当然，模拟程序中的某个参数会发生变化，否则，模拟程序将会以完全相同的方式再运行一次。也许有些数字人希望从过去的某个点开始重新运行（虚拟）现实，并添加上它们上一次未掌握的某些知识。由于整个模拟世界将重新运行，所以它相当于（虚拟）现实的复制品，会耗费相应的计算资源，模拟世界中的所有数字实体也都会被复制。这将引发其他一些问题，但它们与本书内容无关，这里就不做讨论了。

我无法预测数字人总的来说倾向于生活在与外部物理世界高度相似的虚拟世界中，还是倾向于生活在更加灵活和不受物理法则约束（几百万年以来，人类在很多方面都受到物理法则的限制）的模拟环境中。如果我们能创造出虚拟环境，就将真正进入科幻小说描述的世界。我认为，我们的能力目前还不足以思考和处理这些问题。

人格手术和其他改进方法

到目前为止，我们一直假设存在某种程序，它可以确保通过上传生物思维创建的数字思维的行为，在本质上与原初的思维没有区别。然而，合成思维的情况有所不同，因为它们是按照完全不同的原则设计的，它们的工作方式可能与人类思维的工作方式也大不相同。神经形态智能系统介于这两者之间。想要确保生物思维的数字副本和原件的行为相同，最显而易见的方法就是尽可能提高副本的逼真水平。我们知道，创建这种副本所需的技术尚不存在，而且可能永远不会出现。不过，即使我们已经具备了拷贝思维的数字副本的技术，它也不足以让我们理解和修饰仿真思维。

对于上传思维，有人可能会设想拷贝是通过某个自动化的过程完成的，而且这个过程可以直接识别大脑的结构和机制，并将它们拷贝到计算机的存储器中，然后计算机就可以仿真大脑内部发生的过程了。理解大脑的工作原理是另外一个难题，而且要复杂得多。大脑中产生特定感觉或记忆的活动模式非常复杂，难以用第一原理来解释。情绪和其他脑功能可能是大脑的复杂放电模式的涌现性质，不太容易理解，更不要说改变或改进了。

随着我们对大脑工作原理的理解不断深入，终有一天思维的主人可以自主修饰、微调、改进和改造思维。我们或许还可以通过操纵大脑的结构和模式，选择性地抹去某些记忆或提高智力水平。但是，这种操作需要对大脑行为有一定程度的理解，而有能力创造数字思维并不一定意味着我们可以自动满足这种要求。

近几十年来，脑科学取得了重大进展，而且很可能会以越来越快的速度继续向前发展。即使我们有可能详细理解控制大脑内部运转的

机制，也必须记住每个人的大脑各不相同。所以，理解通用机制并不等同于理解特定大脑的行为。一般的化学失衡和导致疾病的其他大规模的过程，可以采用与现代医疗技术十分类似的方法予以解决。想要实现这一目标，只要根据需要改变的环境变量（在现实生活中，这些变量是通过药物、手术或其他宏观尺度的过程控制的），改变模拟程序中的某些参数即可。不过，即使我们可以通过精细的神经元层级的手术消除某段记忆或某种行为，也需要满足一个先决条件，即非常详细地理解每个大脑的行为，这可能比理解支撑大脑行为的通用机制难度要大得多。

有一个因素可以让我们更容易地实现对特定思维行为的细致理解。在观察数字思维的行为时，我们可以达到所需的任意精细程度，但在观察生物思维的行为时，仪器和物理限制会导致我们的观察严重受限。我在第9章讨论了仪器的进步，但即便如此，我们仍然无法获得关于工作大脑中单个神经元放电的详细信息。通过在计算机上仿真人脑而创建的数字思维，可用于检测工作大脑中的每一种放电模式、每一种连接和每一种现象。利用这些信息，研究人员既可以理解现有思维的工作原理，还可以学习如何设计更强大的思维。这就引出了一种有趣的可能性：数字思维或许可以自我改进，因为它们在检测和理解它们自己的行为方面，至少原则上可以达到前所未有的详细程度。

第12章

关于未来的种种猜想

虽然前面章节提出的各种可能性看似遥不可及，但它们说明了一个事实，那就是未来的技术发展可能比人们想象的更加难以预测。在最后一章中，我将讨论技术进化的几个可能的长期方向。这些猜想都是有根据的，当然，你也可以说我在胡思乱想。

技术奇点

自举法这个词可能产生于几百年前，原意是指一种想象的依靠自己的力量将自己提举起来的能力。在这里，它指一个进程一旦开始就具有自维持和持续自增强的能力。

与生物思维相比，数字思维有若干优势。它们可以拷贝任意多次，可以按不同的速度运行（取决于可用的计算能力），还可以用不适合生物思维的方法装配和改造。我们有理由认为，数字思维一旦出现，就会提升我们对智能的理解。单凭这些简单的优势，数字思维就足以在高级人工智能技术的发展中占据重要地位。

在计算机中运行的智能系统可以自引导，在不需要人为干预的情况下变得越来越智能，这一基本理念已经有些年头了，曾以不同的形式出现在科学著作和科幻小说中。有时它被称为种子人工智能，意思是指，只需要创造出第一个足够聪明并且能够理解和改进自身的人工智能系统，就可以创建出一个不断加速的过程。尽管这样的系统尚未出现，甚至未被列入计划，但我们有理由认为，如果数字思维真的出现，那么它将在推动催生它的那些技术进一步发展方面发挥重要作用。

欧文·约翰·古德是一位数学家，曾与阿兰·图灵在布莱切利园共事，为英国破解德国的恩尼格玛密码做出了贡献。1965年，古德创造了“智能爆炸”一词：

> 超智能机器可以被定义为智能水平远超人类所有智力活动的机器。由于设计机器属于这些智力活动之一，所以超智能机器可以设计出更好的机器。毫无疑问，超智能机器将引发“智能爆炸”，人

类的智能将被远远抛在后面。因此，如无必要，人类绝对不应该发明第一台超智能机器，除非这台机器足够温顺并告诉我们如何控制它。

古德认为，即使机器的智能水平只是略微超过人类，它们也肯定能以人类设计师无法预见的方式改进自己的设计，不断提升自己的智能水平，从而远超人类。

自举（通过自举，智能系统将产生智能水平更高的系统，并形成不断加速和不断自增强的循环）的可能性，是促使许多人相信我们最终将看到社会组织方式呈现不连续性的因素之一。这种不连续性被称为技术奇点，或者简称奇点。它是指技术进化会带来速度更快的技术进化，从而形成一个迅速加速的循环，最终促使整个社会发生突然和彻底的变化。人们认为智能机器将在这个循环中发挥重要作用，因为它们可用于设计越来越智能的机器，促使机器的智能水平和技术能力迅速超过人类。由于技术奇点将导致人类丧失在地球上的智能优势地位，所以它出现之后的技术和社会进化将变得完全不可预测。

数学上的奇点指的是，使函数或其他数学对象达到某一特殊值或未定义值，抑或失去某种特定光滑性的点。例如，函数$1/x$在$x=0$时有一个奇点，因为在这一点上函数没有定义值；当x趋近0且取正值时，函数值会迅速趋于无穷大。

受数学奇点概念的启发，人们使用技术奇点来表示人类的技术进化过程丧失连续性的那个时点。第一个使用奇点描述这种现象的人可能是数学家和物理学家冯·诺伊曼。1958年，斯塔尼斯拉夫·乌拉姆在悼念冯·诺依曼时提到了一次对话，冯·诺依曼在这次对话中说，“日益加快的技术进步和人类生活方式的改变，使人类社会正在逼近某个极其重要的奇点。一旦它出现，我们已知的人类事务就再也无法继续”。

将技术奇点这个表述普及开来的人是科学家、科幻作家弗诺·文奇和雷·库兹韦尔。文奇认为，奇点的出现是其他几个原因导致的或推动的。库兹韦尔写作了好几部书，探讨即将到来的奇点。

值得注意的是，除自然与合成的数字思维以外，其他技术也可能导致技术奇点。生物学和医学领域的颠覆性进步，可能会使人类拥有置换或再生身体和大脑的能力，从而使永生或接近永生成为现实。这种情况肯定会破坏许多社会约定，但在我看来这算不上奇点，因为社会大体上会平稳地适应死亡不再不可避免的情况。纳米技术的显著进步，使利用原子创造任意物理对象成为可能，也将改变许多基本的社会规则，从而形成一个任何人都可以获取任何对象的社会。不过，立法和其他限制可能会确保足够的连续性，使我们知道的人类生活仍然是我们可以理解的。

在我看来，这些技术的潜在颠覆性都比不上数字思维的出现。生活在数字世界里的数字人，只受可用计算机能力的限制。它们将创造出一个完全不同于现代社会的社会，“奇点”一词的出现也就顺理成章了。尽可能地提升计算能力，将产生新的计算技术和体系结构，它们也许会调配地球上的大部分资源去创建越来越强大的计算机。科幻小说和未来学家设想整个地球将变成一台计算机，他们甚至想象出一种叫作“计算素”的物质，用该物质制造计算机跟利用现有资源制造计算机同样高效。一些人把计算素定义为一种能从可用的物质和能量中提取尽可能多的计算能力的物质，这种材料将被用于建造终极计算机，以便满足虚拟世界中日益增长的文明需求。事实上，计算机的发展使计算设备的体积越来越小，而功能越来越强大，这似乎正在把我们带向计算素的方向，尽管目前地球上只有一小部分物质和能量被用于制造和操作计算机。不过，晶体管变得越来越小，而且有越来越多的资源被用于计算，这不是恰恰表明了我们正在朝着计算素的方向稳步前进吗？

奇点是否会出现的问题尚未达成共识，更不要说奇点出现的时间了。乐观主义者认为，它将在21世纪上半叶的某个时候到来，可以及时地帮助今天的一些人摆脱衰老导致的不可避免的死亡命运。另一些不那么乐观的人估计，奇点几百年后才会出现。2008年，美国《电气和电子工程师学会会刊》推出了一期以奇点为主题的特刊。许多科学家、工程师、预言家和科幻作家都参与讨论了这个话题，但并未得出明确的结论。

关于奇点这个概念，很多人都提出了反对意见，我们已经遇到了一些显而易见或许也是最有说服力的观点。它们的基本理念是，思维不可能是计算过程的结果，意识和真正的智能只能源于人脑的运转。根据这些观点，由计算机运行产生的其他智能必定是一种次要智能，无法创建奇点出现所需的自加速和自增强的过程。

还有一些反对观点建立在这样的事实基础之上，那就是对未来技术的预测——水下城市、核聚变、星际旅行、飞行汽车等——大多没有实现。所以，他们认为关于技术奇点的预测跟其他许多预测一样，都不可能实现。原因仅仅在于，技术将沿其他路径朝其他方向发展。

很多人反对奇点的概念，主要是基于社会和经济方面的考虑。他们认为，随着越来越多的工作岗位消失并被智能体和机器人取代，人类社会创建更多智能机器的动机将越来越弱。严重的失业问题和消费需求的下降，将摧毁人们创造奇点到来所需技术的动机。贾雷德·戴蒙德在2005年出版的《崩溃：社会如何选择成败兴亡》一书中指出，当超出其所在环境的可持续承载能力时，文化最终会走向崩溃。果真如此的话，我们的文明就不可能会朝着奇点的方向进一步进化。

还有一种观点认为，技术创新的速度现在已经停止上升，事实上正在下降。计算机时钟的速度并没有以同样的速度增长，而且有越来越多的证据表明，物理限制最终将使摩尔定律失效，这些事实都表明我们的创新能力在下降。如果这个观点是正确的，而且创新能力确实

在下降，那么20世纪下半叶和21世纪前几十年里的技术变革速度，应该是它有史以来的最快速度。

关于奇点是否会出现，我没有最终的个人意见。尽管我认为技术发展趋势的本质是指数增长，但我认识到这样的分析只适用于大致情况。如果对技术创新的实际情况进行详细分析，就肯定会发现它与指数趋势存在很大的偏差。事实上，指数函数没有奇点。随着时间的推移，它们增长得越来越快，但它们永远不会达到数学上的奇点。

不过，总体而言，指数发展的速度非常快，但最终会达到某个物理极限，这是一个事实。活细胞的繁殖能力引发的指数增长，只在物理资源充足的情况下才能持续；当达到某个超高水平后，增长速度就会减缓。我们近几十年来观察到的技术的“指数”发展可能会达到极限，之后创新速度会因为技术或社会原因而下降，整个社会以一种更加稳定和可预测的创新方式向前发展。

不过，即使技术奇点不会出现，我仍然认为社会将继续高速发展，到21世纪末，它的变化幅度之大会让生活在今天的人不敢相信自己的眼睛。我在前文中提到过克拉克第三定律，它指出任何足够先进的技术都与魔法无异，飞机、汽车、计算机和手机对生活在几百年前的人来说都是魔法。未来会产生什么样的奇妙技术？答案不仅仅是数字思维！

奇点之后，是生是死？

真正相信奇点会到来的人可能希望采取某些措施，确保他们的人格（如果身体不可以）可以长久保存，以便利用奇点带来的可能性。相信人体冷冻法的人使用的最显而易见的方法是人体超低温保存法。超低温保存是一种成熟技术，在这个过程中，细胞和组织被冷却到极低的温度，以实现长期保存的目的。如果温度足够低（接近80开氏度，差不多就是液氮的温度），可能对生物组织造成损害的酶或化学活动就会被有效地终止，这使延长标本的保存时间成为可能。人体冷冻法的理念是，让人类的大脑或整个身体保持冷冻状态，等待未来相关技术的出现。

人体冷冻法面临的挑战是，既要达到超低温，又要避免这个过程中的某些现象对细胞造成巨大损害。其中破坏性最大的现象是细胞脱水和细胞内外结冰，虽然使用冷冻保护剂之类的化学物质可以减少这些有害影响，但目前的技术仍然会对超过一定大小的器官造成严重损害。

一旦物质被冷冻保存，受到进一步损害的可能性就比较小，而且保存时间可长达数百年。不过，我们可能还需要考虑其他影响，比如辐射引发的细胞损伤。

在现有技术条件下，人体或整个大脑的超低温保存被视为不可逆性操作，因为组织会受到巨大损害。几十年前，人们在尝试恢复已冷冻数千年之久的大型哺乳动物时，就已经舍弃直接增温的方法了。但是，相信人体冷冻法的人想当然地认为，未来的技术可以修复冷冻过程造成的损害，超低温保存的人（或大脑）终有一天会复活。

包括人体冷冻研究所和阿尔科生命延续基金会在内的一些机构及公司，为其会员提供用液氮冷冻保存头部或身体的不同选择，最低报价不少于10万美元。考虑到有较大可能性复活，这个价格怎么看都是非常合理的。根据公布的统计数字，这些机构到今天为止已拥有几千名会员，并且用液氮保存了几百个人的身体。

关于人体冷冻法的可行性，一直存在着激烈的争论。有几位著名科学家在一封支持这项技术的可行性的公开信上签名，但大多数见多识广的人都高度怀疑人体冷冻法是一个不诚实的计划，只是为了从不知情的客户那里骗取钱财。

其他相信奇点会到来的人则持更加乐观的态度，他们希望奇点能及时到来，拯救今天还活着的人。在这些人中，最直言不讳的一个可能是雷·库兹韦尔。他认为，医学和生物学的进步将使今天的人有可能在他们的有生之年经历奇点。

考虑到利用冰冻大脑进行思维上传或思维修复所需技术的复杂性和巨大的技术突破，我个人认为，上述两种期望可能都是错误的。即使人体冷冻法不会对脑组织造成太大的损伤，未来的技术可能也无法利用这个活体大脑的信息，而在我看来，如果思维上传真能实现，这些信息将必不可少。既然我们认为奇点在21世纪末之前不会出现，那么“它能准时到来，使今天的人摆脱死亡宿命”的希望，就必然建立在医疗水平不断提升，从而使人们的预期寿命每过一年就增加一岁的假设之上。这毫无疑问是一种太过乐观的期望！

超级智能的风险

正如我们看到的那样，奇点是否会出现在不久的将来，这个问题目前还没有定论。但是，即使奇点不会发生，超人类智能仍然可能出于人为或意外原因而出现。欧文·约翰·古德的智能爆炸理念即使不会创造出技术奇点，也可能会产生智能水平远超人类的系统。

以阿兰·图灵为代表的人工智能研究人员，曾多次期待人类水平的人工智能会在某一天成真。但是，我们没有什么特别的理由认为，人类表现出的智能水平有什么独到之处。虽然人类确实比其他动物（包括那些大脑体积更大的动物）更聪明，但我们不能就此理所应当地认为人类的智力是智能可以达到的最高水平。

获得超人类智能的方法可能有很多，包括：大幅加快类人推理的速度（比如，全脑仿真器的运行速度可以达到实时速度的10万倍），将大量协同的自然智能和人工智能集中起来，为人类水平的智能提供大量数据和存储器，开发一些目前未知的新智能形式，等等。这些超人类智能之于人类智能，就好比人类智能之于黑猩猩智能。与其他动物一样，黑猩猩的生存现在更多地取决于人类（地球上占支配地位的生命形式），而不是它们自己。这并不是因为我们更强大或数量更多，而是因为我们更智能，并且拥有十分高级的技术。

如果超人类智能真的被开发出来，我们会不会像现在的黑猩猩受制于人类那样，受制于这些超人类智能呢？人类种族的生存有一天也许会取决于超级智能对人类的仁慈之心。既然这种超人类智能是我们直接或间接创造出来的，那么我们也许有机会做一些安排，使它只为人类的最大利益服务。但是，这说起来容易做起来却很难。

1942年，艾萨克·阿西莫夫提出了“机器人三定律”，如果所有机器人都能遵循它们，就可以保证人类的生命安全。机器人三定律是指，第一，机器人不得伤害人类，也不得见到人类受到伤害而袖手旁观；第二，机器人必须服从人类的命令，但不得违反第一条定律；第三，机器人必须保护自身的安全，但不得违反第一条和第二条定律。但是，这些定律现在看起来有些幼稚，因为它们基于这样一种假设（这在人工智能研究早期很常见）：一套明确的符号规则最终将产生强大的人工智能。随着技术的发展，我们现在知道人工智能系统无法通过巨细靡遗的编程来实现，我们也不能通过一套规则全面定义它的行为。人工智能系统只能通过晦涩难懂的学习算法、统计分析和复杂的目标函数，形成它们自己的行为规则。

超人类智能很容易失控，即使它们要实现的目标是人类为其设定的。为便于说明，我们假设有一个超级智能系统被要求解决全球变暖问题。如果这个超级智能系统认定最有效的解决方案是将人类文明带回到技术时代之前的状态，甚至是让人类从地球上消失，事情可能就不会朝着它的设计者期望的方向发展了。你也许认为这种可能性非常小，但事实上，真正的超级智能系统可能会持有与人类截然不同的目标、价值观和方法。真正的超级智能系统可能比人类聪明得多，开发技术和寻找解决方案的效率也很高，以至于人类将沦为其实现目标和手段的工具。此外，它的动机可能和我们完全不一样，尤其当它是一种合成智能的时候，合成智能的行为与人类智能的行为迥然不同。

智能爆炸可能在较短的时间内就会结束，这进一步增加了这个问题的复杂性。人工智能已经发展了几十年，而思维上传技术的发展可能需要不少于50年的时间。因此我们认为，如果智能爆炸真的发生了，它应该会持续几十年。但是，事实可能并非如此。根据前文中的定义，种子人工智能的自我改进速度将远超人工智能技术的发展速度。通过使用大量的计算资源，并以多种方式加快计算速度，人类水平的人工智能系统有可能在几天、几个小时甚至几秒钟内大幅提高它的智能水平。这完全取决于我们无法预知的因素。

一些作者认为，我们在开发可能产生超人类智能的技术之前，应该确保我们能控制它们，并能引导它们朝着造福人类的目标前进。在《超级智能》一书中，尼克·博斯特罗姆深入讨论了这些问题，包括潜在的风险和回报，并提出可以通过若干方法最终提高我们控制和引导这种超人类智能的可能性。但在我看来，我们仍然不清楚如何有效地阻止真正的超人类智能设定它们自己的目标和行动，以避免它们走上与人类利益相悖的道路。

数字思维将变成人类的仆人还是主人？只有未来知道答案。

外星文明在哪里？

宇宙由超过1 000亿个星系组成。在我们的银河系中，大概有2 000多亿颗恒星。如果银河系是一个有代表性的星系，那么宇宙中的恒星总数至少为10^{22}颗。即使按照最低的估计，宇宙中也有超过10^{21}颗恒星，用卡尔·萨根的话说，这个数字远大于地球上的沙粒数量。我们现在知道，很多恒星都有围绕它们旋转的行星，其中又有很多行星可能具备我们已知的支持生命存在的条件。多年来，人类一直想知道，宇宙中是否还有和我们一样的存在。很难想象，如此广阔的地方现在只居住着一个智慧物种——智人。

利用德雷克方程（1961年由射电天文学家弗兰克·德雷克提出，目的是引发寻找地外智慧物的相关科学讨论），我们可以很容易地计算出银河系中具有远程通信能力的地外文明的估计数量：

$$N = R^{*} \times f_{p} \times n_{e} \times f_{l} \times f_{i} \times f_{c} \times L。$$

其中，N表示银河系中可探测到电磁辐射的文明数量，R*表示适合智慧生命发展的恒星形成率，f_p表示拥有行星系统的恒星占比，n_e表示每个类太阳系中适合生命栖居的行星数量，f_l表示适合生命栖居的行星中有生命出现的比例，f_i表示在有生命存在的行星中出现智慧生命的比例，f_c表示地外文明中拥有技术条件并向太空发射可探测其存在的信号的比例，L表示向太空发射可探测信号的文明持续的时间长度。其中一些因子，比如恒星形成率和拥有行星的恒星比例，可以根据巡天观测和系外行星研究较为容易地估算出来，但有些因子则很难甚至不可能准确地估算出来。

德雷克方程中有三个因子尤其难以精确估计。第一个是，在适合生命存在的行星中真的诞生生命的行星占比，目前没有什么可靠的方法可用于估计这个数字，尽管很多研究人员认为条件适于生存的行星很有可能诞生生命。这种观点的部分依据是，在具备合适的条件后不久（从地质学的角度看）地球上就出现了生命，所以支持生命存在的行星中有很多最终会出现生命。不过，人择原理在某种程度上驳斥了这个观点。毕竟，我们对地球上生命出现的历史分析是非常片面的，因为只有那些诞生了生命的行星才有可能支持智慧生命进行这种分析。

第二个难以估算的因子是，在有生命的行星中真正出现智慧生命的比例。有人认为生命最终肯定会发展出智力作为有效的生存工具，因此他们提出这个因子的估计值接近1。而与之相反的观点则认为这个值肯定很小，因为地球上可能有超过10亿个物种，但其中只有一个物种进化出智能。这种观点并不完全正确，因为人属中有多个物种，他们可能都相当聪明，但除了一个物种以外，其他都已经灭绝了。所以，这个因子的正确值仍然存在相当大的不确定性，它可以是0到1之间的任何值。卡莱布·沙夫在他那本迷人的著作《如果，哥白尼错了》中明确指出，由于人类居住在唯一已知的进化出智能生命的行星上，所以很难准确和公正地估算出这两个因子的值。

第三个难以估计的因子是，向太空发射通信信号的文明持续的时间长度。在估算技术文明的持续时间时，我们甚至没有一个例子可供参考，也没有任何物理学原理作为指导。我们无法知道有助于我们探索外太空的科技文明将持续100年还是1亿年。在历史上，有一些文明存在了几十年到几百年不等，但鉴于20世纪发生的技术变革，很难说历史能给我们好的指导意见。事实上，我认为我们还不具备预测未来的技术文明能持续几百万年的适当工具。截至目前我们的技术文明才持续了几百年，仅凭我们的这一经验，不可能合理地推断出未来的技术文明的持续时间。

如果我们可以对在条件合适的行星上诞生生命的可能性给出一个合理的假设，并且赋予f_i（有生命存在的行星上出现智慧生命的比例）一个保守但又不过分悲观的值，比如1%，决定银河系中存在的技术文明数量的重要因子其实就是L，即技术文明持续的时间长度。如果我们假定它的值只有几百年，那么银河系中很可能只有少数几个技术文明，甚至只有一个。如果技术文明可以持续数百万年，那么银河系中可能有几十万个这样的文明。

这就引出了一个关于地外文明的难题，即费米悖论。它们在哪里？恩里科·费米在1950年提出了这个问题。既然银河系中存在很多先进的地外文明，为什么我们却从未看到它们存在的证据？

想要回答这个问题，我们只能基于“地球殊异假说”，或者基于“在有生命的行星上出现智慧生命的情况非常罕见”的观点，或者基于“能向太空发射通信信号的文明只能持续很短时间”的观点。这三种观点都认同，德雷克方程的难以估算的因子中至少有一个值非常小。

地球殊异假说指出，生命的起源和生物复杂性的进化所需的事件和环境组合的出现概率极低，很可能在银河系中只出现过一次，或者少数几次。这一假说导致n_e和f_l（分别表示每个类太阳系中适合生命存在的行星数量和这些行星中真正诞生生命的比例）的取值非常小。如果地球殊异假说是正确的，那么银河系中只有少数行星（甚至只有一个）诞生了生命。与地球相似的行星可能非常罕见，原因有很多：行星轨道和太阳系的长期不稳定状态，相当频繁的行星激变，在一个特定行星上存在维系生命所需的所有东西的可能性很低，等等。

我在前文中介绍了关于f_i（有生命存在的行星上出现智慧生命的比例）的一些讨论。虽然费米悖论的答案可能隐藏在有生命的行星很多但有智慧生命的行星很少的事实之中，但大多数人认为这种可能性

很小，因为它意味着宇宙中遍布生命，但智慧生命只在地球上进化产生。

第三种观点认为文明的持续时间往往不长，以至于来不及相互通信。这种解释可能与本书的中心论题有关。为什么可以与外界通信和发展太空旅行的技术文明只能持续几百年？原因有两个。

一个可能的原因是，这样的文明要么会耗尽行星的资源，要么会导致无法处理的紧张局势，要么会开发出一种敌对的超级智能，以致最终走向自我毁灭。只在我们认为文明的崩溃会导致创造该文明的物种灭绝或者永久性地回到技术时代之前的状态时，这种解释才成立。鉴于我们对人类历史的了解，这种事件似乎不太可能发生，除非它是由一种前所未有的状况引发的，比如，创造了一种对太空通信不感兴趣的敌对超级智能。

另一个原因是，这样的文明并没有灭绝，而是进化到一种更先进的状态，从此不再与外部世界通信，也不再发展太空旅行。矛盾的是，数字思维可以解释费米悖论。所有足够先进的文明最终可能都会创造出丰富多彩的虚拟现实，并创建它们自己的内部合成宇宙，于是它们不再认为太空旅行有必要或有价值。这类文明在数字计算机或其他先进的计算支持系统中创建的虚拟现实，可能异常丰富且极其强大，以至于它们对星际旅行这种最吸引人的探索活动都失去了兴趣。

如果这种解释成立，那么银河系中可能确实存在许多技术文明，它们栖居于功能极其强大的计算机运行的虚拟现实，蜂拥的集体思维忙着创造它们自己的宇宙和物理定律。事实上，我们甚至也有可能生活在这样的虚拟现实中，只不过我们把它叫作宇宙。

致谢

感谢所有支持并帮助我完成本书创作的人，其中包括给我灵感的多位作者，还有我的家人：我的女儿安娜·索菲亚、海伦娜、玛丽安娜和克莱拉，我的父母阿林多和阿玛莉亚。我希望他们有一天能阅读本书，理解其中的一些较为激进的观点。但是，我担心我的父母可能做不到。鉴于科学和技术的性质，我们很快都会落后于时代，无法理解新进展。有时事物变化太快，让我们目不暇接。

我也要感谢我有幸工作过的那些机构，包括里斯本大学、加州大学伯克利分校、铿腾电子科技有限公司、欧洲核子研究组织和葡萄牙国立计算机与系统工程研究所。几十年来，它们就是我的第二个家。

有人说，一本书永远也写不完，除非它的作者决定终止创作。在我从一次非常重要的外科手术中康复后，我为本书的写作画上了一个句号。正是因为我的身体在手术后需要时间恢复，本书才有成书的可能，否则它就只是我的一个梦想。它的诞生应归功于帮助过我的许多朋友，以及为我做手术的优秀外科医生多明戈斯·科伊蒂罗。

本书的整体架构和文本，在很大程度上受益于我的许多朋友和同事，他们耐心地阅读了书稿。我还要特别感谢乔纳斯·阿尔梅达、米格尔·博托、路易斯·佩德罗·柯艾略、蒂亚戈·多明戈斯、马里奥·菲格雷多、亚历山大·弗朗西斯科、维托·莱涛、佩德罗·格德斯·德奥利维拉、阿米尔卡·塞尔纳达斯、马努埃尔·梅代罗·席尔瓦、拉里·瓦尔德，以及为我提供了许多有用建议和修改意见的匿名评论者。我也要感谢米格尔·布加略、阿丰索·迪亚斯、帕特里夏·

菲格雷多、亚历山大·弗朗西斯科、鲍比·卡斯图里和理查德·惠勒，他们为本书提供了许多素材。

感谢佩德罗·多明戈斯将麻省理工学院出版社推荐给我，感谢玛丽·拉夫金·李对本书抱持的信心，感谢凯瑟琳·汉斯莱解决了所有与书稿付梓有关的技术问题。我要特别感谢英文版的文字编辑保罗·贝斯格，他的建议、修改和完善极大地提高了本书的可读性。

在本书写作期间以及之前的几年里，我的妻子安娜·特蕾莎是我的精神支柱和力量源泉，她的付出显然超出了她的职责所在。我衷心地感谢她，并把这本书献给我们共同的未来。

延伸阅读

第1章

本章讨论的主题涉及近代和更早些时候的技术史。对此感兴趣的读者可以参阅尤瓦尔·赫拉利的《人类简史》和贾雷德·戴蒙德的《枪炮、病菌与钢铁》。

第2章

约翰·古德斯布洛姆的《火与文明》和理查德·兰厄姆的《生火》（*Cotching Fire*），生动地描述了史前人类使用火的情况。凯文·凯利的《科技想要什么》从几个方面讨论了技术的发展和未来，与我的观点不谋而合。威廉·罗森的《世界上最强大的创意》（*The Most Powerful Idea in the World*）入木三分地讲述了工业革命带来的颠覆性变化。詹姆斯·格雷克的《信息简史》解释了信息成为现代生活的主导因素的诸多原因。《虚构集》中收录了豪尔赫·路易斯·博尔赫斯的短篇小说《巴别图书馆》。

第3章

在众多讲述20世纪的科学故事的书籍中，我推荐大家阅读罗伯特·奥特的《万物之理》（*The Theory of Almost Everything*）、比尔·布莱森的《万物简史》、斯蒂芬·霍金的《时间简史》，以及阿米尔·埃泽尔的《纠缠》（*Entanglement*）。布鲁斯·亨特的《麦克斯

韦学派》（*The Maxwellians*）详细介绍了麦克斯韦的追随者是如何将麦克斯韦的观点发扬光大的。迈克尔·赖尔登与莉莲·霍德森合著的《晶体之火》（*Crystal Fire*）详细讲述了晶体管是如何改变世界的。理查德·费曼的《QED：光和物质的奇妙理论》可能仍然是最容易理解的量子物理学入门读物。约翰·巴罗的《大自然的常数》对编码宇宙秘密的众多数字进行了生动有趣的解释。罗伯特·拉基的《硅之梦》（*Silicon Dreams*）深入讨论了开创信息时代的那些技术。赫伯特·托布与唐纳德·席林合著的《数字集成电子学》（*Digital Integrated Electronics*），对那些想了解如何设计数字电路的人来说仍然是一本有用的参考书。约翰·亨尼斯与戴维·帕特森合著的《计算机体系结构》是计算机体系结构方面的权威著作。

第4章

伊恩·沃森的《通用机器》（*The Universal Machine*）是一本介绍计算机历史和发展前景的优秀著作。乔治·戴森的《图灵的大教堂》则为我们提供了另一个视角。迈克尔·西普塞在他的《计算理论导引》中向读者介绍了语言、图灵机和计算的微妙之处。托马斯·科尔曼、查尔斯·雷瑟森、罗纳德·李维斯特和克利福德·斯坦合著的《算法导论》中关于算法和数据结构的论述仍然具有参考价值。迈克尔·加里和戴维·约翰逊合著的《计算机与难解性》（*Computers and Intractability*）是从事复杂性研究的人不可或缺的工具书。丹·古斯菲尔德的《字符串、树和序列的算法》（*Algorthms on Strings, Trees, and Sequences*）是一本关于字符串操作算法的综合性学术著作，适合研究生阅读。

第5章

斯图尔特·罗素和彼得·诺维格合著的《人工智能》至今仍是人工智能领域的必备参考书，尽管该领域的文献资料非常多。伊莱恩·里奇的《人工智能》虽然已经过时了，但它提出的用符号方法研究人工智能问题的观点仍然值得我们借鉴。马文·明斯基和西蒙·派珀特合著的《感知器》，对单个感知器的作用和局限性的论述具有不容置疑的权威性和历史意义。戴维·鲁姆哈特、詹姆斯·麦克兰德和PDP研究小组合作完成的《并行分布式处理》（*Parallel Distributed Processing*）是人工智能的现代连接机制的起源，一直具有不容忽视的影响力。佩德罗·多明戈斯在他的《终极算法》中用通俗易懂的语言对机器学习做了生动有趣的介绍。另外，特里瓦·哈斯蒂、罗伯特·提布施瓦尼与杰罗姆·弗里德曼合著的《统计学习基础》，汤姆·米切尔的《机器学习》，朱迪亚·珀尔的《智能系统的概率推理》（*Probabilistic Reasoning in Intelligent Systems*），以及理查德·萨顿与安德鲁·巴托合著的《强化学习》（*Reinforcement Learning*），都对这个主题进行了技术性更强的讨论。

第6章

我对进化问题的看法受到很多著作的影响，其中包括理查德·道金斯的《自私的基因》、《盲人钟表匠》和《祖先的故事》，以及斯蒂芬·杰·古尔德的《熊猫的拇指》。《基因组：人种自传23章》和《先天后天》都是马特·里德利的著作，对基因控制我们生命的方法进行了精彩绝伦的描述。杰弗里·库珀的《细胞分子探索》对细胞组织的研究颇具参考价值。丹尼尔·丹尼特的《达尔文的危险思想》可能是对本书影响最深的几部著作之一。

第7章

讲述人类基因组竞赛故事的书有很多，我们或许可以从詹姆斯·沃森的《DNA：生命的秘密》和克雷格·温特的《解码生命》入手，了解相关往事。克雷格·温特的《光速人生》介绍了合成生物学的诞生。此外，凯文·戴维斯在《序列》中以更加中立的立场介绍了这场竞赛。要了解生物信息技术对当今社会的影响，可以阅读格林·穆迪的《生命的数字密码》（*The Digital Code of Life*）。皮埃尔·巴尔迪与索恩·布鲁纳克合著的《生物信息学》，尼尔·琼斯与帕维尔·帕夫纳合著的《生物信息学算法导论》等书从技术的角度对生物信息学进行了介绍。

第8章

本章主要受到了戴维·休伯尔的《眼睛、大脑和视觉》（*Eye, Brain,and Vision*）和戴维·马尔的《视觉》（*Vision*）这两本书的影响。伯蒂尔·希勒的《可兴奋膜的离子通道》（*Ionic Channels of Excitable Membranes*）仍然是关于神经元膜行为的重要参考资料。斯科特·凯尔索的《动态模式：大脑和行为的自组织》（*Dynamic Patterns: The Self-Organization of Brain and Behavior*）是试图找出大脑的一般性组织原则的少数著作之一。史蒂芬·平克的《心智探奇》和斯坦尼斯拉斯·迪昂的《大脑与意识》，是解释大脑运行原理的有趣尝试，尽管我们在这方面的知识十分贫乏。

第9章

加来道雄的《心灵的未来》一书对大脑研究的诸多方向进行了通俗易懂、生动有趣的描述。史蒂文·罗斯的《记忆的形成》（*The Making of Memory*）和承现峻的《连接组》分别属于经典论著和近期作品，讲述了研究人员为理解大脑的工作原理做出的努力。奥拉夫·

斯鲍恩斯的《大脑的网络》（*Networks of the Brain*）详细介绍了大脑网络的研究方法。

第10章

对关注大脑和思维问题的人来说，侯世达的《哥德尔·艾舍尔·巴赫：集异璧之大成》，罗杰·彭罗斯的《皇帝新脑》，侯世达与丹尼尔·丹尼特合著的《心我论》，以及丹尼特的《脑力劳动的产物》（*Brainchildren*）都是必不可少的参考文献。关于意识和自由意志的著作有很多，我推荐大家阅读苏珊·布莱克摩尔的《意识简介》（*Consciousness: A Very Short Introduction*）、陶·诺瑞钱德的《使用者的错觉》、丹尼尔·丹尼特的《自由的进化》和《意识的解释》、罗伯特·奥恩斯坦的《意识的进化》（*The Evolution of Consciousness*），以及萨姆·哈里斯的《自由意志》（*Free Will*）。埃里克·德雷克斯勒的经典著作《创造的引擎》仍然是纳米技术方面的重要参考文献。许峰雄的《“深蓝”背后》充分讲述了研究人员为创造出打败国际象棋冠军的计算机而付出的努力。格雷格·伊根的《赞蒂奇》给我们营造了一种逼真的体验，以至于模糊了科幻作品和现实之间的界线。

第11章

本章讨论的许多道德问题，在上面提到的丹尼尔·丹尼特的几部著作和他的《头脑风暴》中都有涉及。尼克·博斯特罗姆在《超级智能》中广泛讨论了本章涉及的经济、社会、哲学和技术问题，资料翔实，发人深省。本章讨论的其他话题在一些科幻小说中也能看到，在这里，我推荐查尔斯·斯特罗斯的《终端渐速》和《玻璃屋》、格雷格·伊根的《置换城市》，以及尼尔·斯蒂芬森的《钻石年代》和《雪崩》。

第12章

在文明的进化、人类的未来和奇点的可能性方面，对我产生影响的著作有：贾雷德·戴蒙德的《崩溃：社会如何选择成败兴亡》，雷·库兹韦尔的《机器之心》和《奇点临近》。彼得·沃德和唐纳德·布朗利在他们合著的《地球殊异》（*Rare Earth*）中提出了宇宙中生命非常罕见的观点。卡莱布·沙夫的《如果，哥白尼错了》在探索地外生命方面可能是最全面的参考文献。

参考文献

Agrawal, Manindra, Neeraj Kayal, and Nitin Saxena. 2004. PRIMES Is in P. Annals of Mathematics 160 (2): 781–793.

Almeida, Luis B. 1989. “Backpropagation in Perceptrons with Feedback.” In Neural Computers, ed. R. Eckmiller and C. von der Malsburg. Springer.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J.Lipman. 1990. “Basic Local Alignment Search Tool.” Journal of Molecular Biology 215(3): 403–410.

Anger, Andreas M., Jean-Paul Armache, Otto Berninghausen, Michael Habeck,Marion Subklewe, Daniel N. Wilson, and Roland Beckmann. 2013. “Structures of the Human and Drosophila 80S Ribosome.” Nature 497 (7447): 80–85.

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. 1944. “Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III.” Journal of Experimental Medicine 79 (2): 137–158.

Azevedo, Frederico A. C., Ludmila R. B. Carvalho, Lea T. Grinberg, José Marcelo Farfel, Renata E. L. Ferretti, Renata E. P. Leite, Roberto Lent, and Suzana HerculanoHouzel. 2009. “Equal Numbers of Neuronal and Nonneuronal Cells Make the

Human Brain an Isometrically Scaled Up Primate Brain." Journal of Comparative Neurology 513 (5): 532–541.

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. 1999. Modern Information Retrieval. ACM Press.

Barrell, B. G., A. T. Bankier, and J. Drouin. 1979. "A Different Genetic Code in Human Mitochondria." Nature 282 (5735): 189–194.

Beggs, John, and Dietmar Plenz. 2003. "Neuronal Avalanches in Neocortical Circuits." Journal of Neuroscience 23 (35): 11167–11177.

Belliveau, J. W., D. N. Kennedy, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff,M. S. Cohen, J. M. Vevea, T. J. Brady, and B. R. Rosen. 1991. "Functional Mapping of the Human Visual Cortex by Magnetic Resonance Imaging." Science 254 (5032):716–719.

Berger, Hans. 1929. Über das Elektrenkephalogramm des Menschen. European Archives of Psychiatry and Clinical Neuroscience 87 (1): 527–570.

Berman, Helen M, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. The Protein Data Bank. Nucleic Acids Research 28 (1): 235–242.

Bianconi, Eva, Allison Piovesan, Federica Facchin, Alina Beraudi, Raffaella Casadei,Flavia Frabetti, Lorenza Vitale, Maria Chiara Pelleri, Simone Tassani, and Francesco Piva. 2013. "An Estimation of the Number of Cells in the Human Body." Annals of Human Biology 40 (6): 463–471.

Bissell, Chris. 2007. Historical Perspectives: The Moniac A Hydromechanical Analog Computer of the 1950s. IEEE Control Systems Magazine 27 (1): 69–74.

Blakemore, Colin, and Grahame F. Cooper. 1970. “Development of the Brain Depends on the Visual Environment.” Nature 228 (5270): 477–478.

Bliss, T. V., and G. L. Collingridge. 1993. “A Synaptic Model of Memory: Long-Term Potentiation in the Hippocampus.” Nature 361 (6407): 31–39.

Bliss, T. V., and Terje Lømo. 1973. “Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anaesthetized Rabbit Following Stimulation of the Perforant Path.” Journal of Physiology 232 (2): 331–356.

Blumer, A., A. Ehrenfeucht, D. Haussler, and M. Warmuth. 1987. “Occam’s Razor.” Information Processing Letters 24 (6): 377–380.

Blumer, A., and R. L. Rivest. 1988. Training a 3-Node Neural Net Is NP-Complete. Advances in Neural Information Processing Systems suppl. 1:494–501.

Boas, David A., Anders M. Dale, and Maria Angela Franceschini. 2004. “Diffuse Optical Imaging of Brain Activation: Approaches to Optimizing Image Sensitivity,Resolution, and Accuracy.” Neuroimage 23 (suppl. 1): S275–S288.

Bock, Davi D., Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson,

et al. 2011. “Network Anatomy and In Vivo Physiology of Visual Cortical Neurons.” Nature 471 (7337): 177–182.

Borůvka, O. 1926. “O Jistém Problému Minimálním.” Praca Moravske Prirodovedecke Spolecnosti, no. 3: 37–58.

Bostrom, Nick. 2014. Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Boyden, Edward S., Feng Zhang, Ernst Bamberg, Georg Nagel, and Karl Deisseroth. 2005. “Millisecond-Timescale, Genetically Targeted Optical Control of Neural Activity.” Nature Neuroscience 8 (9): 1263–1268.

Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. 1984. Classification and Regression Trees. CRC Press.

Brin, S., and L. Page. 1998. “The Anatomy of a Large-Scale Hypertextual Web Search Engine.” In Proceedings of the Seventh International World Wide Web Conference. Elsevier.

Brocks, Jochen J., Graham A. Logan, Roger Buick, and Roger E. Summons. 1999. “Archean Molecular Fossils and the Early Rise of Eukaryotes.” Science 285 (5430): 1033–1036.

Brodmann, K. 1909. Vergleichende Lokalisationslehre der Groshirnrinde. Barth.

Brownell, Gordon L., and William H. Sweet. 1953. “Localization of Brain Tumors with Positron Emitters.” Nucleonics 11 (11): 40–45.

Bugalho, Miguel M. F., and Arlindo L. Oliveira. 2008. “An Evaluation of the Impact of Side Chain Positioning on

the Accuracy of Discrete Models of Protein Structures.” Advances in Bioinformatics and Computational Biology 5167: 23 – 34.

Caton, Richard. 1875. The Electric Currents of the Brain. British Medical Journal 2:278 – 278.

Charlesworth, Brian, and Deborah Charlesworth. 2009. “Darwin and Genetics.” Genetics 183 (3): 757 – 766.

Church, Alonzo. 1936. “An Unsolvable Problem of Elementary Number Theory.” American Journal of Mathematics 58 (2): 345 – 363.

Clancy, B., R. B. Darlington, and B. L. Finlay. 2001. “Translating Developmental Time across Mammalian Species.” Neuroscience 105 (1): 7 – 17.

Clarke, Arthur C. 1968. A Space Odyssey. Hutchinson.

Cohen, David. 1968. “Magnetoencephalography: Evidence of Magnetic Fields Produced by Alpha-Rhythm Currents.” Science 161 (3843): 784 – 786.

Compeau, Phillip E. C., Pavel A. Pevzner, and Glenn Tesler. 2011. “How to Apply de Bruijn Graphs to Genome Assembly.” Nature Biotechnology 29 (11): 987 – 991.

Cook, S. A. 1971. “The Complexity of Theorem-Proving Procedures.” In Proceedings of the Third Annual ACM Symposium on Theory of Computing. ACM.

Cooke, S. F., and T. V. Bliss. 2006. Plasticity in the Human Central Nervous System. Brain 129 (7): 1659 – 1673.

Cooper, Geoffrey M., and Robert E. Hausman. 2000. The Cell. Sinauer.

Craik, Kenneth James Williams. 1967. The Nature of Explanation. Cambridge University Press.

Crick, Francis, Leslie Barnett, Sydney Brenner, and Richard J. Watts-Tobin. 1961. "General Nature of the Genetic Code for Proteins." Nature 192 (4809): 1227–1232.

Crombie, Alistair Cameron. 1959. Medieval and Early Modern Science., volume II. Harvard University Press.

Crosby, Peter A., Christopher N. Daly, David K. Money, James F. Patrick, Peter M. Seligman, and Janusz A. Kuzma. 1985. Cochlear Implant System for an Auditory Prosthesis. US Patent 4532930.

Crosson, Bruce, Anastasia Ford, Keith M. McGregor, Marcus Meinzer, Sergey Cheshkov, Xiufeng Li, Delaina Walker-Batson, and Richard W. Briggs. 2010. "Functional Imaging and Related Techniques: An Introduction for Rehabilitation Researchers." Journal of Rehabilitation Research and Development 47 (2): vii–xxxiv.

Darwin, Charles. 1859. On the Origin of Species by Means of Natural Selection. Murray.

Davidson, Donald. 1987. Knowing One's Own Mind. Proceedings and Addresses of the American Philosophical Association 60 (3): 441–458.

Davis, Martin, Yuri Matiyasevich, and Julia Robinson. 1976. Hilbert's Tenth Problem: Diophantine Equations:

Positive Aspects of a Negative Solution. Proceedings of Symposia in Pure Mathematics 28:323–378.

Davis, Martin, Hilary Putnam, and Julia Robinson. 1961. The Decision Problem for Exponential Diophantine Equations. Annals of Mathematics 74 (3): 425–436.

Dawkins, Richard. 1976. The Selfish Gene. Oxford University Press.

Dawkins, Richard. 1986. The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design. Norton.

Dawkins, Richard. 2010. The Ancestor’s Tale: A Pilgrimage to the Dawn of Life. Orion. Denk, Winfried, and Heinz Horstmann. 2004. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. PLoS Biology 2(11): e329.

Dennett, Daniel C. 1991. Consciousness Explained. Little, Brown.

Dennett, Daniel C. 1995. Darwin’s Dangerous Idea. Simon and Schuster.

Dennett, Daniel C. 2003. Freedom Evolves. Penguin.

Descartes, René. 1641. Meditationes de Prima Philosophia.

Diamond, Jared. 1997. Guns, Germs, and Steel. Norton.

Diamond, Jared. 2005. Collapse: How Societies Choose to Fail or Succeed. Penguin.

Dias, A., M. Bianciardi, S. Nunes, R. Abreu, J. Rodrigues, L. M. Silveira, L. L. Wald,and P. Figueiredo. 2015. “A New Hierarchical Brain Parcellation Method Based on Discrete Morse Theory for Functional MRI Data.” In IEEE 12th International Symposium on Biomedical Imaging.

Dijkstra, Edsger W. 1959. “A Note on Two Problems in Connexion with Graphs.”Numerische Mathematik 1 (1): 269 – 271.

Dobelle, William H., M. G.Mladejovsky, and J. P. Girvin. 1974. “Artificial Vision for the Blind: Electrical Stimulation of Visual Cortex Offers Hope for a Functional Prosthesis.” Science 183 (4123): 440 – 444.

Domingos, Pedro. 1999. “The Role of Occam’s Razor in Knowledge Discovery.” Data Mining and Knowledge Discovery 3 (4): 409 – 425.

Domingos, Pedro. 2015. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books.

Domingos, Pedro, and Michael Pazzani. 1997. “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss.” Machine Learning 29 (2 – 3): 103 – 130.

Drexler, K. Eric. 1986. Engines of Creation: The Coming Era of Nanotechnology. Anchor. Drubach, Daniel. 2000. The Brain Explained. Prentice-Hall.

Edmondson, James C., and Mary E. Hatten. 1987. “Glial-Guided Granule Neuron Migration in Vitro: A High-Resolution

Time-Lapse Video Microscopic Study.” Journal of Neuroscience 7 (6): 1928 – 1934.

Egan, Greg. 2010a. Permutation City. Hachette.

Egan, Greg. 2010b. Zendegi. Hachette.

Einstein, Albert. 1905a. Über die von der Molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik 17 (8): 549 – 560.

Einstein, Albert. 1905b. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. Annalen der Physik 17 (6): 132 – 148.

Einstein, Albert. 1905c. Zur Elektrodynamik bewegter Körper. Annalen der Physik 17(10): 891 – 921.

Einstein, Albert, Boris Podolsky, and Nathan Rosen. 1935. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” Physical Review 47(10): 777 – 780.

Feynman, Richard Phillips. 1985. QED: The Strange Theory of Light and Matter. Princeton University Press.

Fiers, Walter, Roland Contreras, Fred Duerinck, Guy Haegeman, Dirk Iserentant,Jozef Merregaert, W. Min Jou, et al. 1976. “Complete Nucleotide Sequence of Bacteriophage MS2 RNA: Primary and Secondary Structure of the Replicase Gene.” Nature 260 (5551): 500 – 507.

Fischl, Bruce, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, et al. 2002. “Whole Brain Segmentation:Automated

Labeling of Neuroanatomical Structures in the Human Brain." Neuron 33(3): 341–355.

Fisher, Ronald A. 1936. "Has Mendel's Work Been Rediscovered?" Annals of Science 1 (2): 115–137.

Fleischner, Herbert. 1990. Eulerian Graphs and Related Topics. Elsevier.

Floyd, Robert W. 1962. Algorithm 97: Shortest Path. Communications of the ACM 5 (6): 345.

Franklin, Allan, A. W. F. Edwards, Daniel J. Fairbanks, and Daniel L. Hartl. 2008. Ending the Mendel-Fisher Controversy. University of Pittsburgh Press.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. The Elements of Statistical Learning. Springer.

Fuller, Richard Buckminster, and John McHale. 1967. World Design Science Decade, 1965–1975. World Resources Inventory.

Garey, Michael R., and David S. Johnson. 1979. Computers and Intractability: A Guide to NP-Completeness. Freeman.

Gaur, Albertine. 1992. A History of Writing. University of Chicago Press.

Glover, Paul, and Sir Peter Mansfield. 2002. "Limits to Magnetic Resonance Microscopy." Reports on Progress in Physics 65 (10): 1489–1511.

Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik 38 (1): 173–198.

Good, Irving John. 1965. “Speculations Concerning the First Ultraintelligent Machine.” Advances in Computers 6 (99): 31–83.

Goudsblom, Johan. 1992. Fire and Civilization. Viking.

Gould, Stephen Jay. 1996. Full House: The Spread of Excellence from Plato to Darwin. Harmony Books.

Gratton, Gabriele, and Monica Fabiani. 2001. “Shedding Light on Brain Function: The Event-Related Optical Signal.” Trends in Cognitive Sciences 5 (8): 357–363.

Grosberg, Richard K., and Richard R. Strathmann. 2007. “The Evolution of Multicellularity: A Minor Major Transition?” Annual Review of Ecology, Evolution, and Systematics 38 (1): 621–654.

Gusfield, Dan. 1997. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press.

Hagmann, Patric, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J. Honey, Van J. Wedeen, and Olaf Sporns. 2008. Mapping the Structural Core of Human Cerebral Cortex. PLoS Biology 6 (7): e159.

Hagmann, Patric, Lisa Jonasson, Philippe Maeder, Jean-Philippe Thiran, Van J. Wedeen, and Reto Meuli. 2006. “Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-Weighted Imaging to Diffusion Tensor Imaging and Beyond 1.” Radiographics 26 (suppl. 1): S205–S223.

Hanson, Robin. 2008. Economics of the Singularity. IEEE Spectrum 45 (6): 45–50.

Harari, Yuval Noah. 2014. Sapiens: A Brief History of Humankind. Harper.

Harmand, Sonia, Jason E. Lewis, Craig S. Feibel, Christopher J. Lepre, Sandrine Prat,Arnaud Lenoble, Xavier Boës, et al. 2015. “3.3-Million-Year-Old Stone Tools from Lomekwi 3, West Turkana, Kenya.” Nature 521 (7552): 310–315.

Hawrylycz, Michael J., Ed S. Lein, Angela L. Guillozet-Bongaarts, Elaine H. Shen,Lydia Ng, Jeremy A. Miller, and Louie N. van de Lagemaat, et al. 2012. “An Anatomically Comprehensive Atlas of the Adult Human Brain Transcriptome.” Nature 489(7416): 391–399.

Hebb, Donald Olding. 1949. The Organization of Behavior: A Neuropsychological Theory. Wiley.

Heckemann, Rolf A., Joseph V. Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. 2006. “Automatic Anatomical Brain MRI Segmentation Combining Label Propagation and Decision Fusion.” NeuroImage 33 (1): 115–126.

Helmstaedter, Moritz, Kevin L. Briggman, Srinivas C. Turaga, Viren Jain, H. Sebastian Seung, and Winfried Denk. 2013. “Connectomic Reconstruction of the Inner Plexiform Layer in the Mouse Retina.” Nature 500 (7461): 168–174.

Herraez, Angel. 2006. “Biomolecules in the Computer: Jmol to the Rescue.” Biochemistry and Molecular Biology Education 34 (4): 255–261.

Higham, Tom, Katerina Douka, Rachel Wood, Christopher Bronk Ramsey, Fiona Brock, Laura Basell, Marta Camps, Alvaro

Arrizabalaga, Javier Baena, and Cecillio Barroso-Ruíz. 2014. “The Timing and Spatiotemporal Patterning of Neanderthal Disappearance.” Nature 512 (7514): 306–309.

Hilbert, David. 1902. “Mathematical Problems.” Bulletin of the American Mathematical Society 8 (10): 437–479.

Hirsch, Helmut V. B., and D. N. Spinelli. 1970. “Visual Experience Modifies Distribution of Horizontally and Vertically Oriented Receptive Fields in Cats.” Science 168(3933): 869–871.

Hobbes, Thomas. 1651. Leviathan: Or, The Matter, Form, and Power of a Commonwealth Ecclesiastical and Civil. Routledge.

Hodgkin, Alan L, and Andrew F. Huxley. 1952. “A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve.” Journal of Physiology 117 (4): 500–544.

Hofstadter, Douglas R., and Daniel C. Dennett. 2006. The Mind’s I: Fantasies and Reflections on Self and Soul. Basic Books.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. 1989. “Multilayer Feedforward Networks Are Universal Approximators.” Neural Networks 2 (5): 359–366.

Hounsfield, Godfrey N. 1973. “Computerized Transverse Axial Scanning(Tomography): Part 1. Description of System.” British Journal of Radiology 46 (552): 1016–1022.

Hsieh, Jiang. 2009. Computed Tomography: Principles, Design, Artifacts, and Recent Advances. Wiley.

Hubel, David H. 1988. Eye, Brain, and Vision. Scientific American Library.

Hubel, David H., and Torsten N. Wiesel. 1962. “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex.” Journal of Physiology 160 (1): 106 – 154.

Hubel, David H., and Torsten N. Wiesel. 1968. “Receptive Fields and Functional Architecture of Monkey Striate Cortex.” Journal of Physiology 195 (1): 215 – 243.

Hume, David. 1748. An Enquiry Concerning Human Understanding. Millar.Hunt, Bruce J. 1991. The Maxwellians. Cornell University Press.

Jaynes, Julian. 1976. The Origin of Consciousness in the Breakdown of the Bicameral Mind. Houghton Mifflin.

Kasthuri, Narayanan, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, et al. 2015. “Saturated Reconstruction of a Volume of Neocortex.” Cell 162 (3): 648 – 661.

Kelly, Kevin. 2010. What Technology Wants. Penguin.

Kelso, J. A. Scott. 1997. Dynamic Patterns: The Self-Organization of Brain and Behavior. MIT Press.

Knight, Tom. 2003. Idempotent Vector Design for Standard Assembly of Biobricks. MIT Artificial Intelligence Laboratory.

Knoll, Andrew H., Emmanuelle J. Javaux, D. Hewitt, and P. Cohen. 2006. “Eukaryotic Organisms in Proterozoic Oceans.” Philosophical Transactions of the Royal Society B: Biological Sciences 361 (1470): 102310–38.

Kononenko, Igor. 1993. Inductive and Bayesian Learning in Medical Diagnosis. Applied Artificial Intelligence 7 (4): 317–337.

Kunkel, S., M. Schmidt, J. Eppler, H. Plesser, G. Masumoto, J. Igarashi, S. Ishii,et al. 2014. Spiking Network Simulation Code for Petascale Computers. Frontiers in Neuroinformatics 8 (10): 1–23.

Kurzweil, Ray. 1990. The Age of Intelligent Machines. MIT Press.

Kurzweil, Ray. 2000. The Age of Spiritual Machines: When Computers Exceed Human Intelligence. Penguin.

Kurzweil, Ray. 2005. The Singularity Is Near: When Humans Transcend Biology. Penguin.

Kurzweil, Ray, and Terry Grossman. 2004. Fantastic Voyage: Live Long Enough to Live Forever. Rodale.

Laird, John E., Allen Newell, and Paul S. Rosenbloom. 1987. “SOAR: An Architecture for General Intelligence.” Artificial Intelligence 33 (1): 1–64.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody,Jennifer Baldwin, Keri Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” Nature 409 (6822): 860–921.

Laureyn, W., C. Liu, C. K. O. Sullivan, V. C. Ozalp, O. Nilsson, C. Fermér, K. S. Drese,et al. 2007. "Lab-on-Chip for the Isolation and Characterization of Circulating Tumor Cells." In Proceedings of the 29th Annual Conference of the IEEE on Engineering in Medicine and Biology Society. IEEE.

Lauterbur, Paul C. 1973. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance." Nature 242 (5394): 190–191.

Le Bihan, Denis, and E Breton. 1985. "Imagerie de Diffusion in-Vivo Par Résonance Magnétique Nucléaire." Comptes-Rendus de l' Académie des Sciences 93 (5): 27–34.

Leitão, A. C., A. P. Francisco, R. Abreu, S. Nunes, J. Rodrigues, P. Figueiredo, L. L.Wald, M. Bianciardi, and L. M. Silveira. 2015. "Techniques for Brain Functional Connectivity Analysis from High Resolution Imaging." In Proceedings of the 6th Workshop on Complex Networks. Springer.

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 1983.Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act. Brain 106 (3): 623–642.

Lilienfeld, Julius. 1930. Method and Apparatus for Controlling Electric Currents. US Patent 1745175.

Lømo, T. 1966. "Frequency Potentiation of Excitatory Synaptic Activity in Dentate Area of Hippocampal Formation." Acta Physiologica Scandinavica 68 (suppl. 277):128.

Lucas, John R. 1961. Minds, Machines and Gödel. Philosophy (London, England) 36(137): 112–127.

Magee, John F. 1964. “How to Use Decision Trees in Capital Investment.” Harvard Business Review 42 (5): 79–96.

Marasco, Addolorata, Alessandro Limongiello, and Michele Migliore. 2013. “Using Strahler’s Analysis to Reduce up to 200-Fold the Run Time of Realistic Neuron Models.” Scientific Reports 3 (2934): 1–7.

Marchant, Jo. 2008. Decoding the Heavens: A 2,000-Year-Old-Computer — and the Century-Long Search to Discover Its Secrets. Heinemann.

Maxwell, James Clerk. 1865. A Dynamical Theory of the Electromagnetic Field. Philosophical Transactions of the Royal Society of London 155:459–512.

Maxwell, James Clerk. 1873. A Treatise on Electricity and Magnetism. Clarendon.

Mazziotta, J., A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, et al. 2001. A Probabilistic Atlas and Reference System for the Human Brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London 356 (1412): 1293–1322.

McCulloch, Warren S., and Walter Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” Bulletin of Mathematical Biophysics 5 (4): 115–133.

McVean, G. A., and 1000 Genomes Project Consortium. 2012. “An Integrated Map of Genetic Variation from 1,092 Human

Genomes.” Nature 491 (7422): 56–65.

Menabrea, Luigi Federico, and Ada Lovelace. 1843. “Sketch of the Analytical Engine Invented by Charles Babbage.” Scientific Memoirs 3: 666–731.

Mendel, Gregor. 1866. “Versuche über Pflanzenhybriden.” Verhandlungen des Naturforschenden Vereines in Brunn 4: 3–47.

Merkle, Ralph C. 1992. “The Technical Feasibility of Cryonics.” Medical Hypotheses 39 (1): 6–16.

Metzker, Michael L. 2010. “Sequencing Technologies — the next Generation.” Nature Reviews Genetics 11 (1): 31–46.

Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell. 2013. Machine Learning: An Artificial Intelligence Approach. Springer.

Miedaner, Terrel. 1978. The Soul of Anna Klane: A Novel. Ballantine Books.

Minsky, Marvin. 1961. “Microscopy Apparatus.” US Patent 3013467.

Minsky, Marvin, and Seymour Papert. 1969. Perceptrons. MIT Press.

Modis, Theodore. 2006. “The Singularity Myth.” Technological Forecasting and Social Change 73 (2): 104–112.

Mondragon, A., S. Subbiah, S. C. Almo, M. Drottar, and S. C. Harrison. 1989. “Structure of the Amino-Terminal Domain

of Phage 434 Represser at 2.0 Å Resolution.” Journal of Molecular Biology 205 (1): 189 – 200.

Moore, Gordon E. 1965. “The Density of Transistors Assembled on a Micro Chip Doubles Every 12 Months.” Electronics 38 (8): 114 – 117.

Myers, Eugene W., Granger G. Sutton, Art L. Delcher, Ian M. Dew, Dan P. Fasulo,Michael J. Flanigan, Saul A Kravitz, et al. 2000. “A Whole-Genome Assembly of Drosophila.” Science 287 (5461): 2196 – 2204.

Nagel, Georg, Martin Brauner, Jana F. Liewald, Nona Adeishvili, Ernst Bamberg, and Alexander Gottschalk. 2005. “Light Activation of Channelrhodopsin-2 in Excitable Cells of Caenorhabditis elegans Triggers Rapid Behavioral Responses.” Current Biology 15 (24): 2279 – 2284.

Nagel, Georg, Tanjef Szellas, Wolfram Huhn, Suneel Kateriya, Nona Adeishvili,Peter Berthold, Doris Ollig, et al. 2003. Channelrhodopsin-2, a Directly Light-Gated Cation-Selective Membrane Channel. Proceedings of the National Academy of Sciences of the United States of America 100 (24): 13940 – 13945.

Nagel, Thomas. 1974. “What Is It Like to Be a Bat?” Philosophical Review 83 (4): 435 – 450.

NCBI (National Center for Biotechnology Information). 2007. The Cells of Eukaryotes and Prokaryotes. https://commons.wikimedia.org/wiki/File:Celltypes.svg.

Newell, Allen, and Herbert A. Simon. 1956. “The Logic Theory Machine — A Complex Information Processing System.”

IRE Transactions on Information Theory 2 (3): 61–79.

Nirenberg, Marshall, and Philip Leder. 1964. “RNA Codewords and Protein Synthesis: The Effect of Trinucleotides upon the Binding of sRNA to Ribosomes.” Science 145 (3639): 1399–1407.

Nørretranders, Tor. 1991. The User Illusion. Penguin.

Ogawa, Seiji, Tso-Ming Lee, Alan R. Kay, and David W. Tank. 1990. Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation. Proceedings of the National Academy of Sciences of the United States of America 87 (24): 9868–9872.

Okada, Y. 1983. “Neurogenesis of Evoked Magnetic Fields.” In Biomagnetism: An Interdisciplinary Approach, ed. S. J. Williamson, G.-L. Romani, L. Kaufman, and I.Modena. Springer.

Olazaran, Mikel. 1996. “A Sociological Study of the Official History of the Perceptrons Controversy.” Social Studies of Science 26 (3): 611–659.

Oliveira, Arlindo L. 1994. Inductive Learning by Selection of Minimal Complexity Representations. PhD dissertation, University of California, Berkeley.

Papert, Seymour. 1966. The Summer Vision Project. http://dspace.mit.edu/handle/1721.1/6125.

Parfit, Derek. 1984. Reasons and Persons. Oxford University Press.

Pauling, Linus. Robert B. Corey, and Herman R. Branson. 1951. “The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain.” Proceedings of the National Academy of Sciences 37 (4): 205–211.

Penrose, Roger. 1989. The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics. Oxford University Press.

Pevzner, Pavel A., Haixu Tang, and Michael S. Waterman. 2001. “An Eulerian Path Approach to DNA Fragment Assembly.” Proceedings of the National Academy of Sciences 98 (17): 9748–9753.

Phillips, A. William. 1950. Mechanical Models in Economic Dynamics. Economics 17 (67): 283–305.

Pires, Ana M., and João A. Branco. 2010. “A Statistical Model to Explain the Mendel–Fisher Controversy.” Statistical Science 25 (4): 545–565.

Post, Emil L. 1946. “A Variant of a Recursively Unsolvable Problem.” Bulletin of the American Mathematical Society 52 (4): 264–268.

Prim, Robert Clay. 1957. “Shortest Connection Networks and Some Generalizations.” Bell System Technical Journal 36 (6): 1389–1401.

Quinlan, J R. 1986. “Induction of Decision Trees.” Machine Learning 1 (1): 81–106.

Ragan, Timothy, Lolahon R. Kadiri, Kannan Umadevi Venkataraju, Karsten Bahlmann, Jason Sutin, Julian Taranda, Ignacio Arganda-Carreras, et al. 2012. Serial Two-Photon

Tomography for Automated Ex Vivo Mouse Brain Imaging. Nature Methods 9 (3): 255–258.

Raichle, Marcus E. 2009. A Paradigm Shift in Functional Brain Imaging. Journal of Neuroscience 29 (41): 12729–12734.

Raichle, Marcus E., and Mark A. Mintun. 2006. "Brain Work and Brain Imaging." Annual Review of Neuroscience 29: 449–476.

Ramón y Cajal, Santiago. 1904. Textura Del Sistema Nervioso Del Hombre Y de Los Vertebrados. Nicolas Moya.

Ramón y Cajal, Santiago. 1901. Recuerdos de mi vida. Fortanet.

Ranson, Stephen Walter. 1920. The Anatomy of the Nervous System. Saunders.

Reimann, Michael W., Costas A. Anastassiou, Rodrigo Perin, Sean L. Hill, Henry Markram, and Christof Koch. 2013. "A Biophysically Detailed Model of Neocortical Local Field Potentials Predicts the Critical Role of Active Membrane Currents." Neuron 79 (2): 375–390.

Rich, Elaine. 1983. Artificial Intelligence. McGraw-Hill.

Riedlinger, Reid, Ron Arnold, Larry Biro, Bill Bowhill, Jason Crop, Kevin Duda,Eric S. Fetzer, et al. 2012. "A 32 Nm, 3.1 Billion Transistor, 12 Wide Issue Itanium®Processor for Mission-Critical Servers." IEEE Journal of Solid-State Circuits 47 (1): 177–193.

Rivest, R. L., A. Shamir, and L. Adleman. 1978. "A Method for Obtaining Digital Signatures and Public-Key

Cryptosystems.” Communications of the ACM 21 (2): 120–126.

Rosenblatt, Frank. 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” Psychological Review 65 (6): 386–408.

Rougier, Nicolas. 2007. Biological Neuron Schema. https://commons.wikimedia.org/wiki/File:Neuron-figure-fr.svg.

Roy, Ambrish, Alper Kucukural, and Yang Zhang. 2010. “I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction.” Nature Protocols 5 (4): 725–738.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning Representations by Back-Propagating Errors. Nature 5 (6088): 533–536.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. 1986. Parallel Distributed Processing, volumes 1 and 2. MIT Press.

Russell, Stuart, and Peter Norvig. 2009. Artificial Intelligence: A Modern Approach. Prentice-Hall.

Samuel, Arthur L. 1959. “Some Studies in Machine Learning Using the Game of Checkers.” IBM Journal of Research and Development 3 (3): 210–229.

Sandberg, Anders, and Nick Bostrom. 2008. Whole Brain Emulation: A Roadmap. Technical Report 2008-3, Future of Humanity Institute.

Sanger, Frederick, Steven Nicklen, and Alan R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.”

Proceedings of the National Academy of Sciences 74 (12): 5463–5467.

Scharf, Caleb. 2014. The Copernicus Complex: Our Cosmic Significance in a Universe of Planets and Probabilities. Scientific American / Farrar, Straus and Giroux.

Schleiden, Matthias Jacob. 1838. "Beiträge zur Phytogenesis." Archiv fur Anatomie,Physiologie und Wissenschaftliche Medicin, 137–176.

Schreyer, Paul. 2000. The Contribution of Information and Communication Technology to Output Growth. Science, Technology and Industry Working Paper 2000/02, OECD.

Schrödinger, Erwin. 1935. "Die gegenwärtige Situation in der Quantenmechanik." Naturwissenschaften 23 (49): 823–828.

Schultz, Thomas. 2006. "DTI-Sagittal-Fibers." https://commons.wikimedia.org/wiki/File:DTI-sagittal-fibers.jpg.

Schuster, Stephan C. 2007. "Next-Generation Sequencing Transforms Today's Biology." Nature Methods 5 (1): 16–18.

Schwann, Théodore. 1839. Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiers und Pflanzen. Reimer.

Searle, John. 1980. Minds, Brains, and Programs. Behavioral and Brain Sciences 3 (03): 417–424.

Searle, John. 2011. "Watson Doesn't Know It Won on 'Jeopardy!'" Wall Street Journal, February 23.

Seung, Sebastian. 2012. Connectome: How the Brain’s Wiring Makes Us Who We Are. Houghton Mifflin Harcourt.

Sheffer, Henry Maurice. 1913. “A Set of Five Independent Postulates for Boolean Algebras, with Application to Logical Constants.” Transactions of the American Mathematical Society 14 (4): 481–488.

Shor, Peter W. 1997. “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer.” SIAM Journal on Computing 26 (5): 1484–1509.

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. 2016. “Mastering the Game of Go with Deep Neural Networks and Tree Search.” Nature 529 (7587): 484–489.

Sipser, Michael. 1997. Introduction to the Theory of Computation. PWS.

Smullyan, Raymond. 1980. This Book Needs No Title: A Budget of Living Paradoxes. Simon and Schuster.

Söding, Johannes. 2005. Protein Homology Detection by HMM–HMM Comparison. Bioinformatics 21 (7): 951–960.

Solomonoff, R. J. 1964. “A Formal Theory of Inductive Inference,” parts I and II. Information and Control 7 (1): 1–22, 7 (2): 224–254.

Sporns, Olaf. 2011. Networks of the Brain. MIT Press.

Stephenson, Neal. 1995. The Diamond Age: Or, A Young Lady’s Illustrated Primer. Bantam Books.

Stiles, Joan, and Terry L. Jernigan. 2010. “The Basics of Brain Development.” Neuropsychology Review 20 (4): 327–348.

Stross, Charles. 2006. Glasshouse. Penguin.

Stuart, Greg, Nelson Spruston, Bert Sakmann, and Michael Häusser. 1997. “Action Potential Initiation and Backpropagation in Neurons of the Mammalian CNS.” Trends in Neurosciences 20 (3): 125–131.

Sutton, Richard S. 1988. “Learning to Predict by the Methods of Temporal Differences.” Machine Learning 3 (1): 9–44.

Sutton, Richard S., and Andrew G. Barto. 1998. Reinforcement Learning: An Introduction. MIT Press.

Turing, Alan. 1937. On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society 2 (42): 230–265.

Turing, Alan. 1950. Computing Machinery and Intelligence. Mind 59 (236): 433–460.

Ulam, Stanislaw. 1958. “Tribute to John von Neumann.” Bulletin of the American Mathematical Society 64 (3): 1–49.

Van Inwagen, Peter. 1983. An Essay on Free Will. Oxford University Press.

Vapnik, Vladimir Naumovich. 1998. Statistical Learning Theory. Wiley.

Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural,Granger G. Sutton, Hamilton O. Smith,

et al. 2001. “The Sequence of the Human Genome.” Science 291 (5507): 1304–1351.

Villringer, Arno, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl. 1993. “Near Infrared Spectroscopy (NIRS): A New Tool to Study Hemodynamic Changes during Activation of Brain Function in Human Adults.” Neuroscience Letters 154 (1): 101–104.

Vinge, Vernor. 1993. “The Coming Technological Singularity: How to Survive in the Post-Human Era.” In Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace. NASA.

von Neumann, John. 1945. First Draft of a Report on the EDVAC. Moore School Library, University of Pennsylvania.

Ward, Peter, and Donald Brownlee. 2000. Rare Earth: Why Complex Life Is Uncommon in the Universe. Copernicus Books.

Watson, James D., and Francis H. C. Crick. 1953. “Molecular Structure of Nucleic Acids.” Nature 171 (4356): 737–738.

Werbos, Paul J. 1988. “Generalization of Backpropagation with Application to a Recurrent Gas Market Model.” Neural Networks 1 (4): 339–356.

Wetterstrand, K. A. 2015. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). National Human Genome Research Institute. http://www.genome.gov/sequencingcosts/.

Wheeler, Richard. 2011. The Structure of DNA Showing with Detail Showing the Structure of the Four Bases, Adenine,

Cytosine, Guanine and Thymine, and the Location of the Major and Minor Groove. https://commons.wikimedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png.

White, J., E. Southgate, J. Thomson, and S. Brenner. 1986. The Structure of the Nervous System of the Nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society of London 314 (1165): 1–340.

Witten, Edward. 1995. "String Theory Dynamics in Various Dimensions." Nuclear Physics B 443 (1): 85–126.

Woese, Carl. 1990. Phylogenetic Tree of Life. Wikimedia Commons.
https://commons.wikimedia.org/wiki/File:PhylogeneticTree.png.

Wolpert, David H. 1996. "The Lack of A Priori Distinctions between Learning Algorithms." Neural Computation 8 (7): 1341–1390.

Wrangham, Richard. 2009. Catching Fire: How Cooking Made Us Human. Basic Books.

Zemelman, Boris V., Georgia A. Lee, Minna Ng, and Gero Miesenböck. 2002.

"Selective Photostimulation of Genetically ChARGed Neurons." Neuron 33 (1): 15–22.